J Philos Logic (2012) 41:347–385
DOI 10.1007/s10992-010-9165-z

Tolerant, Classical, Strict

Pablo Cobreros · Paul Egré · David Ripley · Robert van Rooij

Received: 19 May 2010 / Accepted: 14 October 2010 / Published online: 20 November 2010

Abstract In this paper we investigate a semantics for first-order logic originally proposed by R. van Rooij to account for the idea that vague predicates are tolerant, that is, for the principle that if x is P, then y should be P whenever y is similar enough to x. The semantics, which makes use of indifference relations to model similarity, rests on the interaction of three notions of truth: the *classical* notion, and two dual notions simultaneously defined in terms of it, which we call *tolerant* truth and *strict* truth. We characterize the space of consequence relations definable in terms of those and discuss the kind of solution this gives to the sorites paradox. We discuss some applications of the framework to the pragmatics and psycholinguistics of vague predicates, in particular regarding judgments about borderline cases.

P. Cobreros (✉)
Department of Philosophy, University of Navarra, 31011 Pamplona, Spain
e-mail: pcobreros@unav.es

P. Egré (✉) · D. Ripley
Institut Jean-Nicod (CNRS-EHESS-ENS), Département d'Etudes Cognitives de l'ENS,
29, rue d'Ulm, 75005, Paris, France
e-mail: paul.egre@ens.fr

D. Ripley (✉)
Department of Philosophy—Old Quad, University of Melbourne, Parkville,
VIC 3010, Australia
e-mail: davewripley@gmail.com

R. van Rooij (✉)
Institute for Logic, Language and Computation, Universiteit van Amsterdam,
P.O. Box 94242, 1090 GE, Amsterdam, The Netherlands
e-mail: R.A.M.vanRooij@uva.nl

Keywords Vagueness · Sorites paradox · Tolerance · Logical consequence · Truth · Non-transitivity · Trivalent logics · Paraconsistent logics

Our aim in this paper is to explore a semantic framework originally proposed by R. van Rooij in [29] in order to deal with the sorites paradox, and intended to formalize the idea that vague predicates are tolerant. Standardly, the idea of tolerance is expressed by means of the following principle: if some individual x is P, and x and y are only imperceptibly different in respects relevant for the application of the predicate P, then y is P as well. In classical logic, the principle of tolerance gives rise to the sorites paradox. Because of that, one influential strand of solutions to the sorites paradox consists in rejecting the principle and substituting weaker principles in its stead. A different approach consists in preserving the tolerance principle itself but appealing to a non-classical logic.

The semantics originally proposed by van Rooij belongs to that second family: it allows us to validate the tolerance principle in its plain form, and it is non-classical. The framework rests on the interaction of three notions of truth for sentences involving vague predicates: the classical notion of truth, a notion of tolerant truth, and a dual notion of strict truth. Because of this, the framework leaves room for many different notions of logical consequence. In his earlier work, van Rooij suggested that the appropriate notion should be neither preservation of classical truth nor preservation of tolerant truth, but instead, following motivations given by Zardini in his work on tolerance (see [34]), a mixed notion, on which we reason from classically-true premises to tolerantly-true conclusions.

As it turns out, however, the standard notions of logical consequence for tolerant truth and strict truth are also interesting *per se*. In particular, they bear an unexpected connection to more familiar many-valued logics: the Logic of Paradox (LP) proposed by Priest in [21], and its dual, so-called "Strong Kleene" logic (K3). Because of this, they cast a new light on these many valued approaches, as applied to vagueness. Furthermore, the distinction between tolerant and strict truth also bears a connection to the frameworks of subvaluationism and supervaluationism that have been proposed to deal with vagueness. Because tolerant truth and strict truth are interdefined, however, the semantics makes distinct predictions, in particular regarding borderline cases, for which classical contradictions are predicted to hold tolerantly, and classical validities to fail strictly.

In Section 1, we start out by rehearsing the main motivations behind van Rooij's semantics for the notion of tolerance, and present some basic features of the semantics, in particular regarding the characterization of borderline cases for vague predicates. In Section 2, we characterize logical truths for the notions of tolerant and strict truth, and establish a natural correspondence between tolerant/strict semantics and two well-known many-valued logics, LP and K3. In Section 3, we enlarge the space of consequence relations and discuss various notions of mixed consequence, in particular van Rooij's consequence

from classical to tolerant truth and its kin, and discuss the application of this framework to the sorites paradox. In Section 4, finally, we close this paper with the discussion of some applications of the semantics to the pragmatics and psycholinguistics of vague predicates, in relation to recent experiments by Alxatib and Pelletier [1], Ripley [23] and Serchuk, Hargreaves and Zach [25]. The focus of that section concerns the predictions of tolerant and strict semantics for borderline cases, and in particular the choice between tolerant and strict interpretations for vague predicates.

1 Tolerant and Strict Semantics

1.1 Tolerance and Indifference

Let us consider a vague predicate such as "tall". The principle of tolerance corresponds to the following intuitive constraint: that if one individual is tall, and this individual is not visibly or relevantly taller or smaller than another individual, then the other is tall as well. Formally, the principle can be stated as follows:

(1) $\forall x \forall y (P(x) \wedge x \sim_P y \rightarrow P(y))$, where P stands for "tall", and $\sim_P$ is the relevant indifference relation (namely not looking to have distinct heights).

Central to the discussion of this principle is the specification of the properties of the indifference relation. Arguably, a relation such as "not looking to have distinct heights" is reflexive and symmetric, but not transitive: a can look to have nearly the same height as b, b can look to have nearly the same height as c, but a and c may look to have distinct heights. In our approach, the non-transitivity of the indifference relation is a central feature of all vague predicates (see [11, 32]). One can think of indifference relations of this kind as defined from what Luce in [17] called semi-order relations (see e.g. [29]). In the case of a predicate like "tall", the semi-order associated with it would be the relation $\succ_P$ such that $x \succ_P y$ expresses that x looks visibly or relevantly taller than y. From a semi-order relation, $x \sim_P y$ is definable as $\neg(x \succ_P y \vee y \succ_P x)$, namely neither of x and y looks significantly taller than the other.

We need not specify the properties of semi-orders in this paper (we refer to [29] for details). All we need to assume is that any vague predicate P comes associated with an appropriate indifference relation $\sim_P$ that is reflexive and symmetric, but possibly non-transitive. Van Rooij's proposal is indeed that the tolerance principle can be validated if the semantics of vague predicates is made sensitive to such indifference relations. On van Rooij's approach, we can say that x is tall *tolerantly* if there exists an individual y such that x is similar to y by way of how tall x looks, and y is tall *classically*.

One way to motivate this semantic conception could be the following: suppose that a subject, John, is given a first task, which is to draw as best as he can a sharp line between the tall and the non-tall individuals in a given set. The two sets thus delineated fix a perfectly classical extension for the predicate

"tall" relative to John's inner model of the situation (namely the sets are disjoint and exhaust the whole domain). Now suppose John can still remember where he drew the line, but is assigned a second task, namely is asked of some arbitrary individual x in the series whether x counts as tall or not, with the permission to adjust or correct his initial judgment. What we are assuming is that if x looks sufficiently similar by way of height to an individual y that was put in the set of tall people, then x may be called "tall" as well by John. This would happen even if x is in fact a member of the set of people declared non-tall in the first task, but is such that a slight shift of the line would have counted x as tall originally. Or to put it differently, this would be a situation in which x and y are on either side of the line drawn by John, but nevertheless look very similar in how tall they look. Importantly, however, John can still declare x not-tall in that case, since x looks also sufficiently similar to an individual that was put in the set of non-tall people (namely to x itself, or to non-tall individuals further off the line). On the other hand, however, if x happens to be sufficiently far off the line, namely if x does not look similar in how tall x looks to *any* of the individuals that have been counted as tall, then x will not count as tolerantly tall.

In brief, the intuition behind van Rooij's understanding of tolerance is that whichever way one were to draw the line between the tall and the not-tall, there should remain room to count as tall individuals that are on the other side of the line, provided they are sufficiently close to the tall ones in respects of how tall they look. Understood this way, tolerance corresponds to the possibility of *coarsening* the extension initially assigned to a predicate. In what follows, we shall spell out this idea more formally. We shall proceed in two main steps: we present a first way of articulating the semantics, and explain why it falls short of capturing the idea of tolerance. We then state the official understanding of the notion of tolerance, and explain why it leads us to introduce a dual concept of strict satisfaction for a predicate.

1.2 Preliminaries: Language and Models

Language The language we are interested in in this paper is the language of first-order logic (for now, without identity). To make the exposition simpler, we furthermore restrict the language to monadic predicate logic, as the extension to n-ary predicates does not pose special problems.

Definition 1 Let $\mathcal{P}$ be a denumerable set of unary predicate symbols, $\mathcal{C}$ be a denumerable set of individual constants, and $\mathcal{V}$ be a denumerable set of individual variables. An atomic formula is of the form $P(\underline{a})$ or $P(x)$, where $P \in \mathcal{P}$ and $\underline{a} \in \mathcal{C}$, $x \in \mathcal{V}$.

Definition 2 Well formed formulae (wff): if ϕ is an atomic formula, it is a wff. If ϕ is a wff, so is $\neg\phi$; if ϕ and ψ are wff, so is $(\phi \wedge \psi)$; if ϕ is a wff, so is $\forall x\phi$.

Everywhere, we assume that disjunction $\vee$ and the conditional $\rightarrow$ are defined classically in terms of negation and conjunction. Likewise, $\exists x\phi$ stands for $\neg\forall x\neg\phi$. Brackets are omitted where no ambiguity would result.

Models Classically, satisfaction of first-order formulae is defined over structures of interpretation consisting of a domain of individuals and an interpretation function. We will be interested in expansions of such structures in which every predicate comes with a relation of indifference or similarity. We thus distinguish two kinds of models:

Definition 3 A *C-model* M is a tuple $\langle D, I\rangle$ such that:

- D is a non-empty domain of individuals
- I is an interpretation function (of the usual classical sort) for the non-logical vocabulary: for a constant $\underline{a}$, $I(\underline{a}) \in D$; for a predicate P, $I(P) \in \{0, 1\}^D$.

When no ambiguity results, we write a for $I(\underline{a})$.

Definition 4 A *T-model* M is a tuple $\langle D, I, \sim\rangle$ such that $\langle D, I\rangle$ is a C-model and $\sim$ is a function that takes any predicate P to a binary relation $\sim_P$ on D. For any P, $\sim_P$ is reflexive and symmetric (but possibly non-transitive).[1]

We define satisfaction for wff in a substitutional manner, assuming that given a C-model or T-model, every individual d of the domain has a name $\underline{d}$.[2] If ϕ is a formula, $\phi[\underline{d}/x]$ is the result of substituting $\underline{d}$ for every free occurrence of x in ϕ. We first define classical truth in this way:

Definition 5 *c*-truth in a model. Let M be either a C-model such that $M = \langle D, I\rangle$, or a T-model such that $M = \langle D, I, \sim\rangle$.

$M \vDash^c P(\underline{a})$ iff $I(P)(a) = 1$.
$M \vDash^c \neg\phi$ iff $M \nvDash^c \phi$.
$M \vDash^c \phi \wedge \psi$ iff $M \vDash^c \phi$ and $M \vDash^c \psi$.
$M \vDash^c \forall x\phi$ iff for every d in D, $M \vDash^c \phi[\underline{d}/x]$.

Definition 6 A formula ϕ is classically valid iff every C-model makes it *c*-true. A formula ϕ is *c*-valid iff every T-model makes it *c*-true.

[1] As mentioned above, for every P, the similarity relation $\sim_P$ is taken to be based on a semi-order $\succ_P$, in particular to ensure that T-models adequately model relations of comparison. Pinkal ([20], p. 315) before us defined a notion of T-model ("K-model with tolerance") that also makes central use of similarity relations for each predicate of the language, but based on a space of precisifications of a partial model (thanks to N. Asher for pointing this out to us). His definition of truth in such models does not validate the tolerance principle, however. As in the case of Kamp's 1981 earlier framework (see footnote 3), however, there appears to be important elements of convergence between his approach and ours, which we hope to clarify in future work.

[2] This is unimportant, but it simplifies exposition. It can be replaced with objectual quantification without any trouble.

Fact 1 *Classical validity and c-validity coincide.*

The proof is immediate, since every C-model can be seen as a reduct structure of the corresponding T-model, every T-model as an expansion of the corresponding C-model, and c-truth does not rest on the properties of $\sim$. A consequence is that in what follows, we will be able to work everywhere with T-models; and we will do so, except when we explicitly specify otherwise.

1.3 Tolerance: First Approximation

1.3.1 The Semantics

Let us define a first approximation of the notion of tolerant satisfaction, which we shall write $\models^{t'}$:

Definition 7 t'-truth. Let M be a T-model:

$M \models^{t'} P(\underline{a})$ iff $\exists d \sim_P a : M \vDash^{c} P(\underline{d})$
$M \models^{t'} \neg\phi$ iff $M \not\models^{t'} \phi$
$M \models^{t'} \phi \wedge \psi$ iff $M \models^{t'} \phi$ and $M \models^{t'} \psi$
$M \models^{t'} \forall x\phi$ iff for all $d \in D : M \models^{t'} \phi[\underline{d}/x]$

Definition 8 *Similarity predicates.* For each intended relation of indifference $\sim_P$ over the model, we assume that there is a binary predicate of the language I_P such that by definition $M \vDash^{c} \underline{a} I_P \underline{b}$ iff $M \models^{t'} \underline{a} I_P \underline{b}$ iff $a \sim_P b$. That is, similarity predicates are classically interpreted, even when the relevant notion of satisfaction is tolerant satisfaction.

This assumption will be maintained for the the other notions of truth we will consider in what follows. Essentially, the assumption implies that similarity relations coming with a vague predicate are crisp and extensionally determinate. This may appear to be in tension with the prospect of accounting for vague predicates, but for the theory we develop here what primarily matters is the non-transitive character of such relations.

1.3.2 Evaluation

The semantics we just defined implements the basic idea we described above, but it has two related shortcomings. Let t'-validity for sentences be defined in the expected way, namely as t'-truth in every T-model. Firstly, the principle of tolerance does not come out as a t'-validity in the semantics. Secondly, negation is defined in a very strong way: to say that the negation of a formula is tolerantly true means that the formula is not tolerantly true. A consequence is that tolerance fails to capture the idea of a uniform coarsening of the semantic value of a formula.

Tolerance It is not the case that $\models^{t'} \forall x \forall y(P(x) \wedge xI_P y \rightarrow P(y))$. Consider a structure M with three elements a, b, c such that $a \sim_P b \sim_P c$ but $a \nsim_P c$, and $I(P) = \{a\}$. Clearly, $M \models^{t'} P(\underline{b})$, but $M \not\models^{t'} P(\underline{c})$. Hence the principle of tolerance is not tolerantly valid on this understanding of tolerance.

Negation Let us write $[\![P]\!]^{c,M} = \{d \in M; M \vDash^{c} P(\underline{d})\}$, and $[\![P]\!]^{t',M} = \{d \in M; M \models^{t'} P(\underline{d})\}$. Let us call $[\![P]\!]^{c,M}$ the classical extension of P in M, and $[\![P]\!]^{t',M}$ its tolerant extension. Clearly, for every atomic predicate P of the language, $[\![P]\!]^{c,M} \subseteq [\![P]\!]^{t',M}$, namely the tolerant extension increases the classical extension of the predicate. Let us write $[\![\neg P]\!]^{c,M} = \{d \in M; M \vDash^{c} \neg P(\underline{d})\}$, and similarly $[\![\neg P]\!]^{t',M} = \{d \in M; M \models^{t'} \neg P(\underline{d})\}$. This time, $[\![\neg P]\!]^{t',M} \subseteq [\![\neg P]\!]^{c,M}$, but the converse is not true. This means that it is not true of arbitrary formulae that their tolerant extension in a model is more inclusive than their classical extension. In order to get a uniform notion of coarsening for formulae, it is necessary to weaken the semantics we have here for negation.

1.4 Tolerant and Strict Semantics

To circumvent both these limitations, van Rooij [29] introduced a second notion of tolerant satisfaction, in terms of a dual notion of strict satisfaction, and of classical satisfaction. We write $M \models^{t} \phi$ and $M \models^{s} \phi$ for tolerant and strict satisfaction respectively. The two notions of satisfaction are defined by simultaneous induction.[3]

1.4.1 The Semantics

Definition 9 *t*-truth and *s*-truth. Let M be a T-model:

$M \models^{t} P(\underline{a})$ iff $\exists d \sim_P a : M \vDash^{c} P(\underline{d})$
$M \models^{t} \neg\phi$ iff $M \not\models^{s} \phi$
$M \models^{t} \phi \wedge \psi$ iff $M \models^{t} \phi$ and $M \models^{t} \psi$
$M \models^{t} \forall x\phi$ iff for all $d \in D : M \models^{t} \phi[\underline{d}/x]$

$M \models^{s} P(\underline{a})$ iff $\forall d \sim_P a : M \vDash^{c} P(\underline{d})$
$M \models^{s} \neg\phi$ iff $M \not\models^{t} \phi$
$M \models^{s} \phi \wedge \psi$ iff $M \models^{s} \phi$ and $M \models^{s} \psi$
$M \models^{s} \forall x\phi$ iff for all $d \in D : M \models^{s} \phi[\underline{d}/x]$

Remark 1 By definition, for every formula ϕ: $M \models^{t} \phi$ iff $M \not\models^{s} \neg\phi$, and $M \models^{s} \phi$ iff $M \not\models^{t} \neg\phi$, so $\models^{s}$ and $\models^{t}$ are duals.

[3] We realized after developing the present account that the clauses given here for atomic satisfaction and negation are quite similar to those explored by Kamp in [15, p. 259]. However, Kamp's treatment of conditionals and quantification, as well as his overall framework, is considerably more complex than what we consider here.

Remark 2 As above, we make the assumption that similarity predicates are crisply interpreted relative to each of the notions of truth we have introduced, that is we have: $M \vDash^c \underline{a} I_P \underline{b}$ iff $M \models^t \underline{a} I_P \underline{b}$ iff $M \models^s \underline{a} I_P \underline{b}$ iff $a \sim_P b$. We make a similar assumption for identity predicates.

As explained, this assumption rules out introducing further indifference relations (of the form $\sim_{I_P}$ or $\sim_=$) for the tolerant and strict interpretation of similarity and identity predicates themselves. We are interested in such an approach, but we will not pursue it in this paper. One of the consequences of this assumption is that borderline cases of a predicate are definite on the present approach. We therefore account only for first-order vagueness here. However, it will be seen that even with such a restriction in place, we can get an elaborate account of the link between vagueness, tolerance and the sorites paradox.

1.4.2 Evaluation

Three specific features of the present semantics can be distinguished.

Tolerance First of all, the semantics makes the principle of tolerance *t*-valid.

Fact 2 *For every atomic predicate P,* $\models^t \forall x \forall y (P(x) \wedge x I_P y \rightarrow P(y))$.

Instead of giving a direct proof of Fact 2, we observe that it directly results from the following stronger property of *t*-validities (and from the reflexivity of $\sim_P$ relations):

Fact 3 *For every atomic predicate P,* $\models^t \forall x \forall y \forall z (P(x) \wedge x I_P y \wedge y I_P z \rightarrow P(z))$.

Proof Note that $M \models^t \phi \rightarrow \psi$ iff if $M \models^s \phi$ then $M \models^t \psi$. Suppose that $M \models^s P(\underline{a}) \wedge \underline{a} I_P \underline{b} \wedge \underline{b} I_P \underline{c}$. Since $M \models^s P(\underline{a})$ and $a \sim_P b$, $M \vDash^c P(\underline{b})$. From $b \sim_P c$, it follows that $M \models^t P(\underline{c})$. □

Thus, *t*-truth ensures that tolerance holds up to two steps along the similarity relation. Two is the maximum number of steps that ensure tolerance, however.[4] Consider, for instance, a T-model M with four elements a, b, c, d such that $I(P) = \{a, b\}$, $a \sim_P b \sim_P c \sim_P d$, and nothing else is related by $\sim_P$, except as required by symmetry and reflexivity. In this model, $M \models^t P(\underline{a})$, but $M \not\models^t P(\underline{d})$. Furthermore, $M \models^t P(\underline{a}) \rightarrow P(\underline{b})$, $M \models^t P(\underline{b}) \rightarrow P(\underline{c})$, and $M \models^t P(\underline{c}) \rightarrow P(\underline{d})$. This means that all premises of a standard sorites

[4]What if we wanted to validate only the 1-step version of tolerance, and not the 2-step version? A possibility is to ask for similarity relations to be reflexive, but not necessarily to be symmetric. Symmetry in our models also implies that every model that has at least one borderline case of P (one element tolerantly P and tolerantly not P) must have at least two such elements. Again, giving up on symmetry would allow us to have models with exactly one borderline case. We shall not explore this possibility further here.

can be tolerantly true, without forcing the conclusion to be tolerantly true. Importantly, this implies that *modus ponens* is not a valid inference, if validity is understood as preservation of t-truth. As we shall discuss in Section 3, however, less radical departures from classical logic are still compatible with the t-validity of the tolerance principle.

Negation That the semantics weakens the meaning of negation can be seen from the new clauses. The previous semantics was such that M tolerantly satisfied $\neg\phi$ provided M did not tolerantly satisfy ϕ. Here, M tolerantly satisfies $\neg\phi$ provided M does not *strictly* satisfy ϕ, which is a weaker requirement. A consequence of this is that: $[\![P]\!]^{c,M} \subseteq [\![P]\!]^{t,M}$, and similarly, $[\![\neg P]\!]^{c,M} \subseteq [\![\neg P]\!]^{t,M}$. This property of coarsening is preserved by conjunction, and therefore transfers to all formulae, as we shall prove in the next section. Conversely, it is easy to see that $[\![P]\!]^{s,M} \subseteq [\![P]\!]^{c,M}$, and similarly, $[\![\neg P]\!]^{s,M} \subseteq [\![\neg P]\!]^{c,M}$. This means that in the same way in which the tolerant extension *coarsens* the classical extension of a predicate, the strict extension *sharpens* it.

Borderlines A third and central feature of the semantics is that it allows us to define what it is to be a borderline case of application of a vague predicate in a natural way. Given a T-model M, borderline cases of the application of a predicate P may be defined as those that fall between the tolerant extension and the strict extension of a predicate:[5]

Definition 10 Let $b(P)^M$, the borderline region of a predicate P, be defined as follows $b(P)^M := [\![P]\!]^{t,M} \setminus [\![P]\!]^{s,M}$.

Equivalently, the borderline area of a predicate P can be defined as the set of cases that are neither strictly P, nor strictly not P. This definition is reminiscent of one standard definition of a borderline case of P: a case that is neither definitely P nor definitely not P. Due to the duality of tolerant and strict truth, borderline cases can also be described on the present account as cases that are both tolerantly P and tolerantly $\neg P$.

[5]See also [7] where a very similar definition of borderlineness is proposed, but in a metric setting, in terms of Voronoi diagrams. More generally, our present definition of borderline cases bears a direct analogy to the definition of the *boundary* of a set in topology. Given a topology, the boundary of a set is defined as the difference between the closure of the set and the interior of that set (see e.g. [18]). Tolerant and strict extensions play the same role relative to the classical extension of a predicate in a T-model as do closure and interior relative to a set given a suitable topology. In our setting, in which the relation $\sim_P$ is possibly non-transitive, we cannot straightforwardly equate the operators $[\![\cdot]\!]^s$ and $[\![\cdot]\!]^t$ with interior and closure operators $\mathcal{I}$ and $\mathcal{C}$, so as to satisfy for every P: $\mathcal{I}([\![P]\!]^{c,M}) = [\![P]\!]^{s,M}$ and $\mathcal{C}([\![P]\!]^{c,M}) = [\![P]\!]^{t,M}$. However, we could get this correspondence rigorously by transforming non-transitive T-models into transitive models (see [9], where the operation is called layering).

The same analogy holds with the notions of inner and outer approximation to a set in the theory of rough sets (see [19]). Usual rough sets require an underlying set with an equivalence relation; if we allow the relation to be only reflexive and symmetric, the approach becomes very close to the present one.

An important consequence of this is that some contradictions can be tolerantly true. Consider, for instance, the same model M. (Recall that M consists of four elements a, b, c, d such that $a \sim_P b \sim_P c \sim_P d$, nothing else is $\sim_P$ related except as required by symmetry and reflexivity, and $[\![P]\!]^{c,M} = \{a, b\}$.) In this model, the two individuals b and c around the cutoff between $[\![P]\!]^{c,M}$ and $[\![\neg P]\!]^{c,M}$ are both tolerantly P and tolerantly $\neg P$. The idea that borderline cases support contradictory responses for a predicate is not new. We find it in paraconsistent approaches to vagueness, in particular in Hyde's subvaluationist treatment [13], and in dialetheist approaches based on Priest's Logic of Paradox (see [24, 31]). Our approach rests on different foundations, but in agreement with LP-based treatments, and unlike in subvaluationism, borderline cases of P tolerantly satisfy the conjunction of P and its negation. In the specified model, for instance, $M \models^t P(\underline{b}) \wedge \neg P(\underline{b})$, and similarly for c.

At this point, we should note that the semantics does not commit us to linking assertion exclusively to t-truth rather than s-truth or even c-truth. Because of that, further work needs to be done before we can evaluate whether the present predictions are welcome or unwelcome. Thus, we defer until Section 4 a discussion of the connection between tolerance, strictness, and facts concerning assertion. In the next section, we first investigate the logical properties of our framework in greater detail.

2 Validities and Entailment: Tolerant and Strict

In this section, we characterize both tolerant and strict validities, and the corresponding notions of logical consequence for each notion, namely preservation of tolerant truth and preservation of strict truth. The first part of the section states some basic lemmas concerning validities. The second part gives us a generalization of those results by means of a correspondence between t-validity and LP-validity, and s-validity and K3-validity. An important caveat: in much of this section we restrict the characterization of validity and entailment to formulae that are free of I_P and identity predicates. We will be explicit about when these special predicates are included (we shall call this the *full vocabulary*) and when they are not (the *restricted vocabulary*). We close the section with a brief comparison between the present framework and the frameworks of subvaluationism and supervaluationism on the one hand, and with more familiar semantics for LP and K3 on the other.

2.1 t-validities and s-validities

Definition 11 A formula ϕ is t-valid ($\vDash^t \phi$) iff for every T-model M, $M \models^t \phi$; it is s-valid ($\vDash^s \phi$) iff for every T-model M, $M \models^s \phi$; and it is c-valid ($\vDash^c \phi$) iff for every T-model M, $M \models^c \phi$.

Definition 12 A formula ϕ is *t*-unsatisfiable ($\phi \vDash^t$) iff no T-model M is such that $M \vDash^t \phi$; it is *s*-unsatisfiable ($\phi \vDash^s$) iff no T-model M is such that $M \vDash^s \phi$; and it is *c*-unsatisfiable ($\phi \vDash^c$) iff no T-model M is such that $M \vDash^c \phi$.

Lemma 1 *For any formula ϕ in the full vocabulary, and any T-model M, $M \vDash^c \phi \Rightarrow M \models^t \phi$, and $M \models^s \phi \Rightarrow M \vDash^c \phi$.*

Proof By simultaneous induction on $\models^t$ and $\models^s$.

- Atomic predication: if $M \vDash^c P(\underline{a})$, then clearly $M \models^t P(\underline{a})$, since $a \sim_P a$. And if $M \models^s P(\underline{a})$, then clearly $M \vDash^c P(\underline{a})$ for the same reason.
- I_P and $=$: We have required already that $M \vDash^c \underline{a} I_P \underline{b}$ iff $M \vDash^t \underline{a} I_P \underline{b}$ iff $M \vDash^s \underline{a} I_P \underline{b}$, and similarly for $=$.
- Negation: if $M \vDash^c \neg\phi$, then $M \nvDash^c \phi$, so by induction hypothesis, $M \nvDash^s \phi$, which implies $M \models^t \neg\phi$. If $M \models^s \neg\phi$, then $M \nvDash^t \phi$, and by induction hypothesis, $M \nvDash^c \phi$, ie $M \vDash^c \neg\phi$.
- Conjunction: if $M \vDash^c \phi \wedge \psi$, then $M \vDash^c \phi$ and $M \vDash^c \psi$, and by induction hypothesis, $M \models^t \phi$ and $M \models^t \psi$, ie $M \models^t \phi \wedge \psi$. The case for $\models^s$ is analogous.
- Universal quantification: if $M \vDash^c \forall x\phi$, then for all d in M, $M \vDash^c \phi[\underline{d}/x]$, and by induction hypothesis, for all d, $M \models^t \phi[\underline{d}/x]$, ie $M \models^t \forall x\phi$. The case for $\models^s$ is analogous. □

Corollary 1 *If $\vDash^c \phi$, then $\models^t \phi$.*

Proof If $\vDash^c \phi$, then for every T-model M, $M \vDash^c \phi$. By Lemma 1, for every M, $M \models^t \phi$, hence $\models^t \phi$. □

Corollary 2 *If $\phi \vDash^c$, then $\phi \vDash^s$.*

Proof If $\phi \vDash^c$, then for every T-model M, $M \nvDash^c \phi$, hence by Lemma 1, every M is such that $M \nvDash^s \phi$; hence, $\phi \vDash^s$. □

Lemma 2 *Let M be a C-model of the form $\langle D, I\rangle$, and $M' = \langle D, I, \sim\rangle$ be the T-model obtained from M by letting $a \sim_P b$ iff $a = b$, for every P. Then for every formula ϕ in the restricted vocabulary, $M \vDash^c \phi$ iff $M' \vDash^c \phi$ iff $M' \models^t \phi$ iff $M' \models^s \phi$.*

Proof Obviously $\sim_P$ is an equivalence relation in this case, hence M' is well-defined. Clearly $M \vDash^c \phi$ iff $M' \vDash^c \phi$. The remainder of the proof is by induction on ϕ: we show that if $M' \vDash^t \phi$ then $M' \vDash^s \phi$, and the rest follows from Lemma 1.

- Atomic case: Suppose $M \vDash^t P(\underline{a})$; then there is a d in M such that $d \sim_P a$ and $M' \vDash^c P(\underline{d})$. But the only d such that $d \sim_P a$ is a itself, so $M' \vDash^c P(\underline{a})$. Likewise, since only a is P-similar to itself, for every $d \sim_P a$, $M' \vDash^c P(\underline{d})$, hence $M' \models^s P(\underline{a})$.

- Negation: Suppose $M' \vDash^t \neg\phi$. Then $M' \nvDash^s \phi$. By the induction hypothesis, $M' \nvDash^t \phi$, and so $M' \vDash^s \neg\phi$.
- Conjunction: immediate for both cases.
- Universal quantification: Suppose $M' \vDash^t \forall x\phi$. Then for all d in D, $M' \vDash^t \phi[\underline{d}/x]$. By the induction hypothesis, for all d in D, $M' \vDash^s \phi[\underline{d}/x]$, so $M' \vDash^s \forall x\phi$. □

We can now strengthen Corollaries 1 and 2 to biconditionals:

Theorem 1 *For every formula ϕ in the restricted vocabulary, $\vDash^c \phi$ iff $\models^t \phi$, and $\phi \vDash^c$ iff $\phi \vDash^s$.*

Proof From Corollary 1, we know that $\vDash^c \phi$ entails $\models^t \phi$. Conversely, if $\nvDash^c \phi$, then it means that there is a C-model M such that $M \nvDash^c \phi$. From Lemma 2, it follows that the T-model M' obtained from M by taking $\sim_P$ to be identity for every P is such that $M' \nvDash^t \phi$. Hence $\nvDash^c \phi$ entails $\nvDash^t \phi$.

Similarly, we know from Corollary 2 that $\phi \vDash^c$ entails $\phi \vDash^s$. To show the converse, suppose that $\phi \nvDash^c$. Then there is a C-model M such that $M \vDash^c \phi$. From Lemma 2, we know that the T-model M' derived from M as the Lemma specifies is such that $M' \vDash^s \phi$. Thus, $\phi \nvDash^s$. □

Despite the affinities between t and s on the one hand and classical logic on the other, there are some striking differences. For example, the set of s-validities in the restricted vocabulary is empty (and dually, every sentence in the restricted vocabulary is t-satisfiable). To establish this, the following lemma more than suffices:

Lemma 3 *There is a T-model M such that for every formula ϕ in the restricted vocabulary, $M \nvDash^s \phi$ and $M \models^t \phi$.*

Proof Let M be a T-model in which every atomic predicate P has a classical extension that is neither empty nor equal to the whole domain. For every pair of elements a and b in the domain of M, and for every predicate P, let $a \sim_P b$. By induction, one can show that for every ϕ, $M \nvDash^s \phi$ and $M \models^t \phi$:

- Atomic case: clearly, for every formula of the form $P(\underline{a})$, $M \models^t P(\underline{a})$, since one can find a d P-similar to a that is classically P. Consequently, $M \vDash^t P(\underline{a})$. Dually, for every a, $M \nvDash^s P(\underline{a})$, since a must be P-similar to some element that is not classically P.
- Negation: if $\phi = \neg\psi$. By induction hypothesis, $M \models^t \psi$, and $M \nvDash^s \psi$. If $M \models^s \phi$, then by definition $M \nvDash^t \psi$, which is a contradiction, so $M \nvDash^s \phi$. Since $M \nvDash^s \psi$, then $M \models^t \neg\psi$; that is, $M \models^t \phi$.
- Conjunction and Universal quantification: both cases are straightforward. □

From the previous lemma, it follows immediately that no formula ϕ in the restricted vocabulary is *s*-true in every T-model M, and every ϕ in the restricted vocabulary is *t*-true in at least one model; hence:

Theorem 2 *No formula ϕ in the restricted vocabulary is s-valid. Every formula ϕ in the restricted vocabulary is t-satisfiable.*

Over the restricted vocabulary, we see that tolerant validities coincide with classical validities, and that strict unsatisfiability coincides with classical unsatisfiability. On the other hand, we can see that no formula is tolerantly unsatisfiable, and that no formula is strictly valid.

Because of that, we can already observe that the logics induced by *s*-truth and *t*-truth do not coincide with supervaluationism [10, 16] nor with subvaluationism [13]. These frameworks associate the language with a set of admissible (classical) precisifications. Then, a sentence is supervaluationistically true if and only if it is classically true in *every* admissible precisification, and it is subvaluationistically true if and only if it is true in *at least one* admissible precisification. Based on these quantification patterns, one may have expected *t*-truth to coincide with sub-truth, and *s*-truth with super-truth. In both sub- and super-valuationism, however, validity for formulas coincides with classical validity; this implies in particular that *s*-validity is distinct from supervaluationist validity.[6] Dually, classical contradictions are not subvaluationistically satisfiable; this implies that *t*-satisfaction does not coincide with subvaluationist satisfaction.

However, *t*-validities and *s*-validities appear to coincide exactly with logical validities in two well-known extensions of the logic FDE of first-degree entailment, namely with Priest's Logic of Paradox on the one hand (LP), and the strong Kleene logic on the other (K3). This coincidence is not fortuitous, as we proceed to show in the next subsection.

2.2 Correspondence with LP and K3

LP and its dual K3 are often presented as three-valued logics, and we present them here in this way. In what follows we use the values 1, 1/2, and 0. The values 1 and 0 can be read as truth and falsity, respectively; 1/2 indicates an intermediate status. Advocates of LP often understand the value 1/2 as applying in the overlap of truth and falsity, and advocates of K3 often understand it as applying in the gap between truth and falsity. For our immediate formal purposes, of course, it doesn't matter how we interpret this value. Entailment in K3 corresponds to preservation of the value 1 from premises to conclusions, and entailment in LP on the other hand corresponds to preservation of nonzero value from premises to conclusions.

[6] For detailed discussions of supervaluationist systems of consequence, see e.g. [4, 5, 30].

If we define *s*-entailment as preservation of strict truth, and *t*-entailment as preservation of tolerant truth, a natural correspondence immediately arises between the two frameworks: we let value 1 represent strict truth, 0 represent strict falsity, and 1/2 represent borderline truth (in the sense we defined in the previous section, see Definition 10). Tolerant truth then corresponds to assigning value 1 or 1/2 to a formula (that is, nonzero value), since a formula is tolerantly true either if it is strictly true, or if it is borderline true.

2.2.1 MV-models and Entailment

To establish the correspondence more formally, we first introduce the notion of an MV-model (for many-valued model). We use MV-models only over the restricted vocabulary, and do not at any point extend them to include I_P or $=$ predicates. We let $\min(A)$ denote the minimum value in the set A, and use $\min(x, y)$, when x and y are numbers, to abbreviate $\min(\{x, y\})$.

Definition 13 An *MV-model* M is a tuple $\langle D, I\rangle$ such that:

- D is a non-empty domain of individuals; and
- I is a three-valued interpretation that works as follows:
  - For any name $\underline{a}$, $I(\underline{a}) \in D$;
  - For any predicate P, $I(P) \in \{1, 1/2, 0\}^D$;
  - $I(P(\underline{a})) = I(P)(I(\underline{a}))$;
  - $I(\neg\phi) = 1 - I(\phi)$;
  - $I(\phi \wedge \psi) = \min(I(\phi), I(\psi))$;
  - $I(\forall x\phi) = \min(\{I(\phi[\underline{d}/x]) : d \in D\})$

Definition 14 An MV-model $M = \langle D, I\rangle$ *LP-satisfies* a wff ϕ ($M \vDash^{LP} \phi$) iff $I(\phi) > 0$. An MV-model $M = \langle D, I\rangle$ *K3-satisfies* a wff ϕ ($M \vDash^{K3} \phi$) iff $I(\phi) = 1$.

We subsume all notions of *c*-consequence, *s*-consequence, *t*-consequence, LP-consequence, and K3-consequence under the following definition (for $X =$ *c*, *s*, or *t*, an X-model is a T-model; for $X =$ LP or K3, an X-model is an MV-model):

Definition 15 For any logic X, let X-consequence be defined as follows: $\Gamma \vDash^X \Delta$ iff for every X-model M such that $M \vDash^X \gamma$ for every $\gamma \in \Gamma$, $M \vDash^X \delta$ for some $\delta \in \Delta$.

For our purposes here, we rely on available axiomatizations of logical consequence in LP and K3. In particular, we refer to [21, 22] or [3] for details. We rehearse some prominent features of these logics: both logics validate the classical rules of conjunction introduction, conjunction elimination, De Morgan laws for conjunction and negation, double negation introduction as well as elimination, universal generalization, and universal instantiation. In K3,

moreover, every formula is entailed by a classical contradiction; in LP, dually, every classical validity is entailed by any formula.

2.2.2 Model Correspondence

To transfer these results to *s*-consequence and *t*-consequence, we show that for every MV-model, we can construct an equivalent T-model, and vice versa.

Definition 16 A T-model M is *equivalent* to an MV-model M' over a set $\mathcal{L}$ of wff iff for every wff $\phi \in \mathcal{L}$:

- $M \vDash^t \phi$ iff $M' \vDash^{LP} \phi$, and
- $M \vDash^s \phi$ iff $M' \vDash^{K3} \phi$

Lemma 4 *Let $\mathcal{L}$ be our language, using only the restricted vocabulary. For every T-model $M = \langle D, I, \sim \rangle$, there is an MV-model equivalent over the language $\mathcal{L}$.*

Proof We define the equivalent MV-model $M' = \langle D', I' \rangle$ as follows:

- $D' = D$
- For any name $\underline{a}$, $I'(\underline{a}) = I(\underline{a})$
- For any predicate P and any $d \in D$:
  - If $M \vDash^s P(\underline{d})$, then $I'(P)(d) = 1$
  - If $M \nvDash^t P(\underline{d})$, then $I'(P)(d) = 0$
  - Otherwise, $I'(P)(d) = 1/2$

  By Lemma 1, we know that $\{\phi : M \vDash^s \phi\} \subseteq \{\phi : M \vDash^t \phi\}$, so the cases here are exclusive and exhaustive.

Now we show that M is equivalent to M', by an induction on formula construction. The base case is immediate. Inductive case:

$\neg$ Suppose $\phi = \neg\psi$, and the inductive hypothesis holds for ψ. Then $M \vDash^t \phi$ iff $M \nvDash^s \psi$ iff $M' \nvDash^{K3} \psi$ iff $I'(\psi) < 1$ iff $I'(\phi) > 0$ iff $M' \vDash^{LP} \phi$. Similarly, $M \vDash^s \phi$ iff $M \nvDash^t \psi$ iff $M' \nvDash^{LP} \psi$ iff $I'(\psi) = 0$ iff $I'(\phi) = 1$ iff $M' \vDash^{K3} \phi$.

$\wedge$ Suppose $\phi = \psi \wedge \chi$, and the inductive hypothesis holds for ψ and χ. Then $M \vDash^t \phi$ iff ($M \vDash^t \psi$ and $M \vDash^t \chi$) iff ($M' \vDash^{LP} \psi$ and $M' \vDash^{LP} \chi$) iff $\min(I'(\psi), I'(\chi)) > 0$ iff $M' \vDash^{LP} \phi$. Similarly, $M \vDash^s \phi$ iff ($M \vDash^s \psi$ and $M \vDash^s \chi$) iff ($M' \vDash^{K3} \psi$ and $M' \vDash^{K3} \chi$) iff $\min(I'(\psi), I'(\chi)) = 1$ iff $M' \vDash^{K3} \phi$.

$\forall$ Suppose $\phi = \forall x \psi$, and the inductive hypothesis holds for $\psi[\underline{d}/x]$ for every d. Then $M \vDash^t \phi$ iff ($M \vDash^t \psi[\underline{d}/x]$ for all $d \in D$) iff ($M' \vDash^{LP} \psi[\underline{d}/x]$ for all $d \in D$) iff $\min(\{I'(\psi[\underline{d}/x]) : d \in D\}) > 0$ iff $M' \vDash^{LP} \phi$. Similarly, $M \vDash^s \phi$ iff ($M \vDash^s \psi[\underline{d}/x]$ for all $d \in D$) iff ($M' \vDash^{K3} \psi[\underline{d}/x]$ for all $d \in D$) iff $\min(\{I'(\psi[\underline{d}/x]) : d \in D\}) = 1$ iff $M' \vDash^{K3} \phi$.

□

Lemma 5 *For every MV-model $M = \langle D, I \rangle$, there is a T-model equivalent over the set of wff in the restricted vocabulary.*

Proof The equivalent T-model $M' = \langle D', I', \sim \rangle$ will operate with an expanded domain.

We define the T-model as follows:

- $D' = \{\langle d, i\rangle : d \in D, i \in \{0, 1\}\}$
- For any name $\underline{a}$ in the old language, $I'(\underline{a}) = \langle I(\underline{a}), 0\rangle$. Add a new name $\underline{a}'$ to the language for every old name $\underline{a}$, and let $I'(\underline{a}') = \langle I(\underline{a}), 1\rangle$.
- For any predicate P and old name $\underline{a}$:
  - $I'(P)(I(\underline{a})) = 1$ iff $I(P)(I(\underline{a})) = 1$, and $I'(P)(I(\underline{a})) = 0$ otherwise;
  - $I'(P)(I(\underline{a}')) = 1$ iff $I(P)(I(\underline{a})) > 0$, and $I'(P)(I(\underline{a}')) = 0$ otherwise;
  - $\sim_P$ contains $\langle I'(\underline{a}), I'(\underline{a})\rangle, \langle I'(\underline{a}), I'(\underline{a}')\rangle, \langle I'(\underline{a}'), I'(\underline{a})\rangle$, and $\langle I'(\underline{a}'), I'(\underline{a}')\rangle$
- For any predicate P, $\sim_P$ contains nothing more than is given for each old name $\underline{a}$ in the last clause

We now show that the models are equivalent over the old language. Proof is by induction on formula construction. The base step is where all the action is.

$M \vDash^{K3} P(\underline{a})$ iff $I(P)(I(\underline{a})) = 1$ iff $I'(P)(I(\underline{a})) = I'(P)(I(\underline{a}')) = 1$ iff ($M' \vDash^c P(\underline{d})$ for all $d \in D$ such that $d \sim_P a$ (since $a \sim_P a$, $a' \sim_P a$, and nothing else $\sim_P a$)) iff $M' \vDash^s Pa$. Similarly, $M \vDash^{LP} P(\underline{a})$ iff $I(P)(I(\underline{a})) > 0$ iff $I'(P)(I(\underline{a}')) = 1$ iff ($M' \vDash^c P(\underline{d})$ for some $d \in D$ such that $d \sim_P a$) iff $M' \vDash^t P(\underline{a})$.

Inductive step is quick:

$\neg$ $M \vDash^{LP} \neg\phi$ iff $I(\phi) < 1$ iff $M \nvDash^{K3} \phi$ iff $M' \nvDash^s \phi$ iff $M' \vDash^t \neg\phi$. Similarly, $M \vDash^{K3} \neg\phi$ iff $I(\phi) = 0$ iff $M \nvDash^{LP} \phi$ iff $M' \nvDash^t \phi$ iff $M' \nvDash^s \neg\phi$.

$\wedge$ $M \vDash^{LP} \phi \wedge \psi$ iff $\min(I(\phi), I(\psi)) > 0$ iff ($M \vDash^{LP} \phi$ and $M \vDash^{LP} \psi$) iff ($M' \vDash^t \phi$ and $M' \vDash^t \psi$) iff $M' \vDash^t \phi \wedge \psi$. Similarly, $M \vDash^{K3} \phi \wedge \psi$ iff $\min(I(\phi), I(\psi)) = 1$ iff ($M \vDash^{K3} \phi$ and $M \vDash^{K3} \psi$) iff ($M' \vDash^s \phi$ and $M' \vDash^s \psi$) iff $M' \vDash^s \phi \wedge \psi$.

$\forall$ $M \vDash^{LP} \forall x\phi$ iff $\min(\{\phi[\underline{d}/x] : d \in D\}) > 0$ iff ($M \vDash^{LP} \phi[\underline{d}/x]$ for all $d \in D$) iff ($M' \vDash^t \phi[\underline{d}/x]$ for all $d \in D'$) iff $M' \vDash^t \forall x\phi$. Similarly, $M \vDash^{K3} \forall x\phi$ iff $\min(\{\phi[\underline{d}/x] : d \in D\}) = 1$ iff ($M \vDash^{K3} \phi[\underline{d}/x]$ for all $d \in D$) iff ($M' \vDash^s \phi[\underline{d}/x]$ for all $d \in D'$) iff $M' \vDash^s \forall x\phi$.

□

Theorem 3 *For all sets of wff* Γ *and* Δ *in the restricted vocabulary,* $\Gamma \vDash^t \Delta$ *iff* $\Gamma \vDash^{LP} \Delta$, *and* $\Gamma \vDash^s \Delta$ *iff* $\Gamma \vDash^{K3} \Delta$.

Proof Immediate. □

2.3 The Full Vocabulary

So far, our discussion has focused mainly on our restricted vocabulary—in which neither identity nor our family of similarity relations is expressible. Here, we consider the effects created by the expansion to our full vocabulary, in which both identity and similarity predicates occur.

Language As before, we use the language of the quantified monadic predicate calculus, including an identity relation $=$, and, for every unary predicate P, a binary relation I_P, which will express the similarity relation $\sim_P$. To ease notation and except when confusion would result, from now on we shall write a instead of $\underline{a}$ for constants, and Pa instead of $P(\underline{a})$.

Identity and similarity are interpreted as described above, in Section 1. Both relations are always interpreted classically; there is no difference between a model's strictly satisfying, classically satisfying, or tolerantly satisfying any sentence built entirely from identity or similarity relations.[7]

2.3.1 Identity

In this section, we consider the effect that introducing identity has on our consequence relations. The first thing to note is that introducing identity breaks the proofs that $\vDash^t = \vDash^{LP}$and $\vDash^s = \vDash^{K3}$. For consider sentences like $(\forall x \forall y(x = y)) \rightarrow (Pa \vee \neg Pa)$. Although this sentence is not valid in K3 (there might be only one thing, and still that thing might satisfy neither P nor its negation), it is strictly satisfied (and therefore both classically and tolerantly satisfied) by every T-model. After all, although $Pa \vee \neg Pa$ can fail to be strictly satisfied in a T-model M, it can only do so when there are two things a and b in M's domain such that $a \sim_P b$, $M \vDash^c Pa$ and $M \vDash^c \neg Pb$. This means a and b must be distinct.

Similarly, although $\forall x \forall y(x = y) \wedge Pa \wedge \neg Pa$ is satisfiable in LP (there might be only one thing, and still that thing might satisfy both P and its negation), it cannot be tolerantly satisfied (and therefore cannot be classically or strictly satisfied) by any T-model. After all, although $Pa \wedge \neg Pa$ can be tolerantly satisfied by a T-model M, it can only do so in precisely the same circumstances as are required for $M \nvDash^s Pa \vee \neg Pa$. And again, that requires two distinct objects in the domain.

So although strict and tolerant consequence are very similiar to K3 and LP consequence, and indeed are the same in the restricted vocabulary, once we expand our vocabulary to include identity we see that they are distinct. This is because, while MV-models allow us to assign nonclassical values directly to atomic predications, T-models allow us to do so only via covert quantification over the domain. If the domain includes only one thing, T-models can no longer provide non-classical values for any atomic predications. Thinking along these lines yields the following:

Fact 4 *If* $\Gamma \vDash^c \Delta$*, then, where* $\vDash^m$ *is either* $\vDash^s$ *or* $\vDash^t$*:*

- $\Gamma \cup \{\forall x \forall y(x = y)\} \vDash^m \Delta$*, and*
- $\Gamma \vDash^m \Delta \cup \{\neg \forall x \forall y(x = y)\}$

[7] If we were to relax these constraints, much of what we are about to claim would fail; we do not know precisely what the resulting systems would look like, although we are interested in pursuing the issue in future work.

Proof Suppose $\Gamma \vDash^c \Delta$, and suppose M is a countermodel for any of the four inferences in the consequent of Fact 4. If M is a countermodel to the first inference, it must (strictly or tolerantly, whichever matters) satisfy $\forall x \forall y(x = y)$, and thus there is only one member in M's domain. If M is a countermodel to the second inference, it must fail to (strictly or tolerantly, again) satisfy $\neg\forall x \forall y(x = y)$; again, there is only one member in M's domain. So no matter which inference M is supposed to be a counterexample to, there is only one member in M's domain.

By examination of the clauses for atomic predication, we can see that this requires that $M \vDash^s Pa$ iff $M \vDash^c Pa$ iff $M \vDash^t Pa$, for any atomic sentence Pa. What's more, we know that identity and similarity predications are satisfied in each of the three ways if in any. Induction on formula complexity shows that, for any sentence ϕ, $M \vDash^s \phi$ iff $M \vDash^c \phi$ iff $M \vDash^t \phi$. We know that it is not the case both that $M \vDash^c \gamma$ for every $\gamma \in \Gamma$ and that $M \nvDash^c \delta$ for every $\delta \in \Delta$ (since $\Gamma \vDash^c \Delta$). But that means it can't be both that $M \vDash^s \gamma$ for every $\gamma \in \Gamma$ and that $M \nvDash^s \delta$ for every $\Delta \in \delta$; nor can it be both that $M \vDash^t \gamma$ for every $\gamma \in \Gamma$ and that $M \nvDash^t \delta$ for every $\delta \in \Delta$. Thus, M is not a counterexample after all to either of the inferences in question. Contradiction. □

Similar effects can be created by restricting the domain in other ways. For example, $\forall x \forall y \forall z(x = y \vee x = z \vee y = z)$, $Pa \wedge \neg Pa \vDash^t Pb \wedge \neg Pb$. Given the first premise (that there are at most two things), the only way for the second premise to be tolerantly satisfied by a model M is if there are two objects in M's domain that bear $\sim_P$ to each other, exactly one of which classically satisfies P. Whichever one of these objects b picks out, $M \vDash^t Pb \wedge \neg Pb$. This argument is not valid in LP, however. For similar reasons, $\forall x \forall y \forall z(x = y \vee x = z \vee y = z)$, $Pa \vee \neg Pa \vDash^s Pb \vee \neg Pb$, but the argument is not K3-valid. Still more validities can be found along these lines: so long as one object in a model's domain tolerantly satisfies $Px \wedge \neg Px$, another object in the domain must as well; and so long as every object but one in a model's domain strictly satisfies $Px \vee \neg Px$, the last one must as well.

2.3.2 Similarity

The biggest difference introduced by similarity has to do with the principle of tolerance for a predicate P: $\forall x \forall y(Px \wedge xI_P y \rightarrow Py)$. This principle, strictly speaking, cannot be stated in LP, since the language of LP does not include our special I_P predicates. We might state an analogue of the principle by adding to LP I_P predicates required to be reflexive and symmetric, but even then the principle would not be LP-valid.[8] However, the principle is tolerantly valid

[8]For a countermodel, consider an LP-model with two members of the domain, a and b. Let I_P be the universal relation on the domain, and let $I(P)(I(\underline{a})) = 1$ and $I(P)(I(\underline{b})) = 0$. This model is an LP-counterexample to tolerance. This is possible because LP has no way to recognize the connection between I_P and P.

on our models, as shown in Section 1. Similarly, although the negation of this tolerance principle is satisfiable in K3-augmented-with-reflexive-symmetric-I_P-predicates, it is not strictly satisfiable on our models (since to be strictly unsatisfiable just is to have a tolerantly-valid negation).

There will be more differences created by these similarity relations when we examine more articulated notions of consequence in Section 3; we shall discuss those differences there.

2.4 Comparisons

The correspondence we established between tolerant semantics and LP on the one hand, and between strict semantics and K3 on the other, is worth commenting on for several reasons.

First of all, as briefly emphasized at the end of Section 2.1, one might have expected *s*-truth and *t*-truth to behave like subvaluationist truth and supervaluationist truth, based on the *prima facie* analogy between the quantification patterns involved in each case. However, what we see is that the semantics make quite distinct predictions, in particular regarding borderline cases. In sub- and super-valuationism, borderline cases are predicted to satisfy classical validities. In particular, every individual is predicted to be tall or not tall, including an individual who is a borderline case of tallness. Conversely, no individual can be both tall and not tall, not even borderline cases. In the present case, by contrast, every individual is *tolerantly* tall or not tall, but some individuals, namely borderline cases, are tolerantly both. By contrast, not all individuals are *strictly* tall or not tall in a model, since borderline cases are predicted to be neither strictly tall, nor strictly not tall. In our view, and as argued by [24] in relation to the application of LP and K3 to vagueness, these specific predictions for borderline cases are not unwelcome. Rather, unlike for sub- and super-valuationism, they imply that a special semantic status is acknowledged of borderline cases.

A second feature of our target semantics is that, while it coincides with the predictions of the many-valued logics LP and K3, it answers to a distinct motivation. Rather than seeing truth as a unified notion to which sentences might answer in three (or more) different ways, our approach posits distinct notions of truth, each of which a sentence may have or fail to have, but none of which is many-valued.

A third feature of the present semantics is that it gives us a psychologically plausible characterization of borderline cases as cases equisimilar with cases that would support opposed categorizations if subjects were forced to be bivalent. This characterization of borderline cases agrees with other accounts based on the notion of similarity.[9] Furthermore, the characterization of borderline

[9]See [7], where borderline cases of color predicates, in particular, are basically characterized as cases equidistant between prototypes in the relevant conceptual space.

cases within T-models allows us to make sense of the idea that borderline cases are shifty or ambivalent cases, namely cases that can be conceptualized under opposed points of view. We shall say more about this in Section 4.2.3 below.

A fourth feature of the present approach concerns the characterization we gave of logical consequence. Above in Section 1.4, we pointed out that *t*-semantics can make the main premise of a sorites true without paradox, but only because modus ponens is no longer a *t*-valid (or LP-valid) rule of inference. Just as in LP, one objection to the present treatment might be that we fail to adequately capture the real meaning of the conditional when we model it in terms of negation and conjunction in the present framework.[10] More generally, the definition we adopted for *t*-entailment may appear to depart too much from classical logic to provide a decent solution to the sorites.

In the next section, however, we examine various alternatives to the definition of tolerant entailment we examined here. Given that we have three notions of truth we can work with, namely tolerant, classical, and strict, there is indeed room for more consequence relations that just preservation of classical truth, preservation of tolerant truth, or preservation of strict truth. At the end of the section, we explain how the present framework allows us to defuse the sorites paradox.

3 Mixed Consequence

Although we have discussed three distinct notions of consequence thus far—strict, classical, and tolerant—our models in fact give us the materials to define a number of additional notions of consequence. Some of these additional notions, we believe, are philosophically interesting in their own right. In fact, we believe that the most natural notion of consequence flowing from these models is not any of the three we have discussed so far, but a mixed notion, in which the standard for truth is higher in the premises than in the conclusion. Arguments in favor of the exploration of such mixed consequence in relation to vagueness and non-transitivity were given and investigated formally by Zardini in [34], and directly inspired the present proposal.

First, a structured way to talk about many different consequence relations:

Definition 17 Where m and n are s, c, or t, and Γ and Δ are sets of formulas: $\Gamma \models^{mn} \Delta$ iff every T-model M such that $M \models^{m} \gamma$ for every $\gamma \in \Gamma$ is also such that $M \models^{n} \delta$ for some $\delta \in \Delta$.

That is, for an argument to be *mn*-valid is for every model that *m*-satisfies all of the premises to *n*-satisfy at least one of the conclusions. From our three

[10] See [24] for discussion of this point on LP.

notions of satisfaction, this immediately generates nine notions of consequence. The first thing to notice is that three of our nine consequence relations have already been characterized. That is because $\models^{tt} = \models^{t}$, $\models^{cc} = \models^{c}$, and $\models^{ss} = \models^{s}$. These unmixed relations have been dealt with earlier in the paper. We also generalize the notions of formula validity and unsatisfiability:

Definition 18 A formula ψ is *mn*-valid ($\models^{mn} \psi$) iff $\emptyset \models^{mn} \psi$. A formula ψ is *mn*-unsatisfiable ($\psi \models^{mn}$) iff $\psi \models^{mn} \emptyset$.

Note that *mn*-validity amounts to *n*-validity (since whatever *m* is, every model *m*-satisfies every member of the empty set), and that *mn*-unsatisfiability amounts to *m*-unsatisfiability (since whatever *n* is, no model *n*-satisfies any member of the empty set).

The purpose of this section is to explore these notions of consequence, characterize the relations between all nine notions, and offer some philosophical reasons for being interested in 'mixed' consequence relations (that is, consequence relations where $m \neq n$). We first define the notion of *duality* relating some of our nine consequence relations, then we compare these relations, attending to their respective strength. In what follows we will say that a consequence relation is *stronger* than another if it is set-theoretically more inclusive, and *weaker* if it is set-theoretically less inclusive. Note that these terms could easily be exchanged, since the weaker a relation of consequence is in that sense, the more stringent are the requirements it actually puts on the derivation of conclusion from premises; conversely, the stronger it is in the sense here specified, the less stringent are the requirements it involves.

3.1 Duality

Definition 19 (Dual consequence relation) Let $\models^{x}$ be a notion of logical consequence. Its *dual* is the notion of logical consequence $\models^{y}$ such that: $\Gamma \models^{x} \Delta$ iff $\neg(\Delta) \models^{y} \neg(\Gamma)$ (where $\neg(\Delta) = \{\neg\delta \mid \delta \in \Delta\}$).

The duality between notions of logical consequence is based on the duality of the notions of satisfaction and the duality of $\forall$ and $\exists$ in the definition of logical consequence. Though this fact is perhaps a bit too obvious to require a proof, we might express the relations of duality in a synthetic way as follows.

Definition 20 *For a notion of satisfaction indexed as x, $d(x) = y$ just in case, for any M, $M \models^{x} \neg\varphi$ iff $M \not\models^{y} \varphi$*

Lemma 6 $\models^{mn}$ *is the dual of* $\models^{d(n)d(m)}$

Proof Assume: $\Gamma \models^{mn} \Delta$, then:

For every M: if $\forall\gamma \in \Gamma$, $M \models^{m} \gamma$ then $\exists\delta \in \Delta$, $M \models^{n} \delta$

iff
For every M: if $\forall\delta \in \Delta$, $M \nvDash^{n} \delta$ then $\exists\gamma \in \Gamma$, $M \nvDash^{m} \gamma$
iff
For every M: if $\forall\delta \in \Delta$, $M \vDash^{d(n)} \neg\delta$ then $\exists\gamma \in \Gamma$, $M \vDash^{d(m)} \neg\gamma$
iff
$\neg(\Delta) \vDash^{d(n)d(m)} \neg(\Gamma)$ □

This yields immediately the duality relations over the possible combinations of consequence relations since we already know that $d(c) = c$, $d(s) = t$ and $d(t) = s$. More particularly:

1. $\vDash^{cc}$, $\vDash^{st}$ and $\vDash^{ts}$ are self-dual.
2. $\vDash^{ss}$ is the dual of $\vDash^{tt}$.
3. $\vDash^{sc}$ is the dual of $\vDash^{ct}$.
4. $\vDash^{cs}$ is the dual of $\vDash^{tc}$.

3.2 Relations Between Consequence Relations

Lemma 7 *For any notion of satisfaction m,* $\vDash^{tm} \subseteq \vDash^{cm} \subseteq \vDash^{sm}$. *For any notion of satisfaction m,* $\vDash^{ms} \subseteq \vDash^{mc} \subseteq \vDash^{mt}$.

Proof Since we know that, for any model M, $\{\phi : M \vDash^{s} \phi\} \subseteq \{\phi : M \vDash^{c} \phi\} \subseteq \{\phi : M \vDash^{t} \phi\}$ (Lemma 1), it follows that if a model M is a *sm*-counterexample to an argument, it is also a *cm*-counterexample, and if it is a *cm*-counterexample, it is also a *tm*-counterexample. Similarly, if a model is a *mt*-counterexample to an argument, it must also be a *mc*-counterexample, and if it is an *mc*-counterexample, it must also be an *ms*-counterexample. □

This lemma answers some questions about the relations between our nine notions, but not all. We go on to complete the picture, first in the restricted vocabulary we used earlier (restricted to exclude both identity and similarity relations), and then in the full vocabulary.

3.3 Restricted Vocabulary

It is useful to keep in mind the following facts:

1. For any T-model M, $M \vDash^{s} \varphi \Rightarrow M \vDash^{c} \varphi \Rightarrow M \vDash^{t} \varphi$ (Lemma 1).
2. For any C-model M, there is a T-model M' s.t. $M' \vDash^{c} \phi$ iff $M' \vDash^{t} \phi$ iff $M' \vDash^{s} \phi$ iff $M \vDash^{c} \phi$ (Lemma 2).
3. Every formula is *t*-satisfiable and no formula is *s*-valid (Lemma 3).

3.3.1 $\models^{st}$, $\models^{cc}$, $\models^{sc}$, and $\models^{ct}$ coincide

Lemma 8 $\Gamma \models^{st} \Delta \Rightarrow \Gamma \models^{cc} \Delta$.

Proof Assume $\Gamma \not\models^{cc} \Delta$, then:

$$\exists M : \forall \gamma \in \Gamma \; M \models^{c} \gamma \text{ and } \forall \delta \in \Delta \; M \not\models^{c} \delta$$

$$\Downarrow$$

$$\exists M' : \forall \gamma \in \Gamma \; M' \models^{s} \gamma \text{ and } \forall \delta \in \Delta \; M' \not\models^{t} \delta$$

□

Since we know from Lemma 7 that $\models^{cc} \subseteq \models^{st}$, this shows that $\models^{cc} = \models^{st}$. Lemma 7 also gives us $\models^{cc} \subseteq \models^{sc} \subseteq \models^{st}$ and $\models^{cc} \subseteq \models^{ct} \subseteq \models^{st}$, so we can conclude that $\models^{cc} = \models^{sc} = \models^{ct} = \models^{st}$.

3.3.2 $\models^{tc}$ is strictly weaker than $\models^{tt}$ and $\models^{cs}$ is strictly weaker than $\models^{ss}$

We know from Lemma 7 that $\models^{tc} \subseteq \models^{tt}$ and that $\models^{cs} \subseteq \models^{ss}$. Now we show that $\Gamma \models^{tt} \Delta \not\Rightarrow \Gamma \models^{tc} \Delta$ and $\Gamma \models^{ss} \Delta \not\Rightarrow \Gamma \models^{cs} \Delta$. In general, $\varphi \models^{tt} \varphi$ and $\varphi \models^{ss} \varphi$ but, for example, $Px \not\models^{tc} Px$ and $Px \not\models^{cs} Px$. After all, we can have a model M such that $M \models^{c} Px$ but $M \not\models^{s} Px$, or such that $M \models^{t} Px$ but $M \not\models^{c} Px$. Thus, the relations $\models^{tc}$ and $\models^{cs}$ are not reflexive.

On the other hand, some instances of reflexivity do hold. For example, $\forall x\, Px \models^{cs} \forall x\, Px$, and $\exists x\, Px \models^{tc} \exists x\, Px$. One must be wary here: $\forall x(Px \vee \neg Px) \not\models^{cs} \forall x(Px \vee \neg Px)$, and $\exists x(Px \wedge \neg Px) \not\models^{tc} \exists x(Px \wedge \neg Px)$. Thus, these logics do not obey uniform substitution, either.

These two logics are, however, transitive.[11] Let us consider $\models^{tc}$. Suppose $\Gamma, \phi \models^{tc} \Delta$ and $\Gamma \models^{tc} \phi, \Delta$. There is no model M such that $M \models^{t} \gamma$ for every $\gamma \in \Gamma \cup \{\phi\}$ and $M \not\models^{c} \delta$ for every $\delta \in \Delta$. Similarly, there is no model M such that $M \models^{t} \gamma$ for every $\gamma \in \Gamma$ and $M \not\models^{c} \delta$ for every $\delta \in \Delta \cup \{\phi\}$. Now, suppose for reductio there is a model M' such that $M' \models^{t} \gamma$ for every $\gamma \in \Gamma$ and $M' \not\models^{c} \delta$ for every $\delta \in \Delta$. Then it must be that $M' \not\models^{t} \phi$, and it must be that $M' \models^{c} \phi$. But this is impossible (by Lemma 1). So there is no such M', and $\Gamma \models^{tc} \Delta$. The proof is the same for $\models^{cs}$, *mutatis mutandis*.

[11] The version of transitivity we consider here—if both $\Gamma, \phi \models \Delta$ and $\Gamma \models \phi, \Delta$ are valid, then $\Gamma \models \Delta$ is valid—is one of many possible variations. In fact, *tc* and *cs* satisfy any version of transitivity we are aware of.

Exactly what inferences hold in these logics? Later, in Section 3.5, we will provide a tableau-based proof theory that is sound and complete for all nine notions of consequence. For now, we hope these remarks are enough to convey the flavor; $\models^{tc}$ and $\models^{cs}$ are odd logics indeed.

3.3.3 $\models^{ts}$ is strictly weaker than both $\models^{tc}$ and $\models^{cs}$

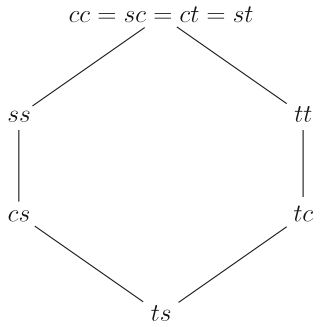
Lemma 9 *$\Gamma \models^{ts} \Delta$ is empty.*

Proof $\Gamma \not\models^{ts} \Delta$ just in case there is a model M such that: $\forall \gamma \in \Gamma$ $M \models^{t} \gamma$ and $\forall \delta \in \Delta$ $M \not\models^{s} \delta$. Now, Lemma 3 shows that there is a model M' in which for every formula φ, $M' \models^{t} \varphi$ and $M' \not\models^{s} \varphi$. Thus, $\Gamma \not\models^{ts} \Delta$ for any Γ and Δ. □

Since both $\models^{tc}$ and $\models^{cs}$ are non-empty, $\models^{ts}$ is strictly weaker than both. (Indeed, $\models^{ts}$—the empty relation—is the weakest possible consequence relation.)

3.3.4 Summing up

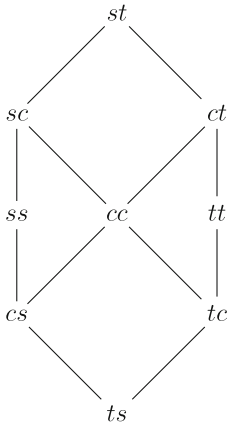
Regarding the nine notions of logical consequence that we might define in the present framework, there are six distinct consequence relations on the restricted vocabulary. $\models^{cc}$, $\models^{ct}$, $\models^{sc}$, and $\models^{st}$ coincide. $\models^{tt}$ and $\models^{ss}$ are distinct and both strictly weaker than $\models^{cc}$.[12] $\models^{tc}$ and $\models^{cs}$ are also distinct, and are strictly weaker than $\models^{tt}$ and $\models^{ss}$ respectively. Finally, $\models^{ts}$ is strictly weaker than both $\models^{tc}$ and $\models^{cs}$.



[12]This can be proved parallel to the above inclusions, or it follows directly from the facts that *tt*-consequence is LP-consequence, *ss*-consequence is K3-consequence, and *cc*-consequence is classical consequence.

3.4 Full Vocabulary

Unlike the restricted vocabulary, in the full vocabulary all nine notions of consequence result in distinct consequence relations, arranged by strength as follows:



We can demonstrate the distinctness of the four strongest relations as follows: the inference from $\{Pa, aI_Pb\}$ to $\{Pb\}$ is *sc*, *ct*, and *st*-valid, but not *cc*-valid. Tolerance—the sentence $\forall x \forall y(Px \wedge xI_Py \rightarrow Py)$—is *ct* and *st*-valid, but not *cc* or *sc*-valid. Tolerance's negation is *sc* and *st*-unsatisfiable, but *cc* and *ct*-satisfiable. What's more, *st*-validity is not simply the union of *sc*- and *ct*-validity, since the inference from $\{Pa, aI_Pb, bI_Pc\}$ to $\{Pc\}$ is *st*-valid, but is neither *sc*- nor *ct*-valid.

Lemma 7, recall, holds for the full vocabulary as well. Further, all the examples that showed distinctness between systems in Section 3.3 still hold. The only remaining question, then, is whether $\models^{tt}$ and $\models^{ss}$ are both still weaker than $\models^{cc}$, or whether the new vocabulary disrupts that relationship. In fact, the new vocabulary does disrupt that relationship: recall that tolerance is *tt*-valid, but not *cc*-valid, and its negation is *ss*-unsatisfiable, but not *cc*-unsatisfiable. So no inclusions hold between $\models^{tt}$, $\models^{ss}$, and $\models^{cc}$ in the full vocabulary.

The deduction theorem (in the form $\Gamma \models^{mn} \Delta$ iff $\models^{mn} \bigwedge \Gamma \rightarrow \bigvee \Delta$)[13] holds for some of our relations, but not all:

Lemma 10 *The deduction theorem* ($\Gamma \models^{mn} \Delta$ *iff* $\models^{mn} \bigwedge \Gamma \rightarrow \bigvee \Delta$) *holds for a consequence relation* $\models^{mn}$ *iff* $m = d(n)$ *(that is, iff* $\models^{mn}$ *is self-dual).*

[13] Since we do not include infinitary conjunction or disjunction in our language, here Γ and Δ must be assumed to be finite sets.

Proof For the right-to-left direction of Lemma 10, assume that $m = d(n)$. Then:

- $\Gamma \models^{mn} \Delta$ iff
- every T-model M is such that either $M \not\models^{m} \bigwedge \Gamma$ or else $M \models^{n} \bigvee \Delta$ iff
- every T-model M is such that either $M \models^{d(m)} \neg \bigwedge \Gamma$ or else $M \models^{n} \bigvee \Delta$ iff
- every T-model M is such that either $M \models^{n} \neg \bigwedge \Gamma$ or else $M \models^{n} \bigvee \Delta$ iff
- every T-model M is such that $M \models^{n} \bigwedge \Gamma \rightarrow \bigvee \Delta$ iff
- $\models^{mn} \bigwedge \Gamma \rightarrow \bigvee \Delta$

For the left-to-right direction: Assume the deduction theorem holds for $\models^{mn}$. Recall that $\models^{mn} \phi$ iff $\models^{xn} \phi$, for any x, ϕ; that is, formula validity depends only on the second notion of satisfaction involved. We know that the deduction theorem holds for $\models^{d(n)n}$, so we can conclude that $\Gamma \models^{mn} \Delta$ iff $\models^{mn} \bigwedge \Gamma \rightarrow \bigvee \Delta$ iff $\models^{d(n)n} \bigwedge \Gamma \rightarrow \bigvee \Delta$ iff $\Gamma \models^{d(n)n} \Delta$. But as we have seen, there are no notions x, m, n of satisfaction such that $\models^{mn} = \models^{xn}$ and $m \neq x$. So $m = d(n)$. □

Thus, the deduction theorem holds for *st*, *cc*, and *ts* only, in the full vocabulary. (In the restricted vocabulary, it holds in addition for *sc* and *ct*, since those are the same as *cc* and *st* in the absence of $=$ and I_P.)

3.4.1 Choosing a Consequence Relation: Modus Ponens, Tolerance, and the Deduction Theorem

Let us see how the stronger notions of consequence here manage to validate tolerance while avoiding soritical reasoning. Consider a soritical sequence of people, arranged by height. Let P be 'tall'; thus I_P will be the usual similarity relation in respect of height. Thus, we have a sequence $\langle a_1, a_2, \ldots, a_n \rangle$, where the first i members are classically P and the remainder are not, and such that $a_i \sim_P a_{i+1}$ for $1 \leq i < n$. Let us focus on *ct*-consequence, for concreteness. We know that Pa_1, $a_1 I_P a_2$, and $a_2 I_P a_3$ all hold, and we know that $Pa_1 \wedge a_1 I_P a_2 \wedge a_2 I_P a_3 \models^{ct} Pa_2 \wedge a_2 I_P a_3$. What's more, we know that $Pa_2 \wedge a_2 I_P a_3 \models^{ct} Pa_3$. It seems we are being led down the soritical series, forced to conclude first Pa_2, then Pa_3, and so on. (We might need to conjoin in more facts about the similarity relation, but nothing should stop us from doing that.) However, this is not the case. Although the above-mentioned inferences do hold in *ct*, $Pa_1 \wedge a_1 I_P a_2 \wedge a_2 I_P a_3 \not\models^{ct} Pa_3$. That is, $\models^{ct}$ is not transitive.[14] So although it

[14]That is, it fails even the simple transitivity principle: if $\phi \models \psi$ and $\psi \models \chi$, then $\phi \models \chi$. There are more general transitivity principles one might consider (with side premises and/or side conclusions, see footnote 11), but violation of this simple transitivity principle suffices for violation of these as well.

validates each step of the soritical reasoning, it does not validate chaining those steps together. *sc* and *st* behave similarly in this regard.[15]

Why would we be interested in nontransitive consequence relations?[16] There are a few reasons. For one thing, we seem to reason nontransitively in the presence of soritical sequences. If we want to capture that reasoning, we must use a nontransitive relation to do the capturing. Additionally, one might think, as we do, that simply using a consequence relation on which tolerance is valid (as it is on $\models^{tt}$) is not quite getting at what we want, if modus ponens isn't valid on that consequence relation (as it is not on $\models^{tt}$). In fact, modus ponens is valid on all four of our strongest relations, since it is *cc*-valid, and anything *cc*-valid is *sc*-, *ct*-, and *st*- valid too. Thus, both *ct*- and *st*-validity are relations on which both tolerance and modus ponens are valid. Such a relation can't reasonably be transitive; it would force us to reason soritically. But a nontransitive relation fits the bill nicely.

Indeed, when it comes to choosing a consequence relation to focus on, we prefer *st* to *sc* on the grounds that tolerance is *st*-valid, and it seems to us that tolerance ought to be valid. We prefer *st* to *ct* because the deduction theorem holds for *st*, and it seems to us that the deduction theorem ought to hold. *st* is the only notion of consequence that satisfies these two desiderata. Thus, we think *st* is the best-motivated of our consequence relations; it validates tolerance, satisfies the deduction theorem, and supports modus ponens. It is quite a reasonable relation for reasoning with vague predicates.

It is worth noting that the semantics of *st* bear similarities to other semantics that have been proposed for consequence relations for vague language. For example, N. Smith in [26] presents an orthodox fuzzy semantics (with sentences taking values from the closed real interval [0,1]) for atomic sentences and the connectives, but defines consequence as follows (notation changed):

> $\Gamma \models \delta$ iff every model on which every $\gamma \in \Gamma$ takes a value strictly greater than .5 is also such that δ takes a value greater than or equal to .5.

Note that this consequence relation sets a stricter standard for its premises than for its conclusion, just as *st* does (and *ct* and *sc* do). However, Smith's consequence relation, unlike ours, is transitive; in fact, it is the consequence relation of classical logic. (Recall that, in the restricted vocabulary, the same is true of *st*, *ct*, and *sc*.)

Another direct relative of the present semantics is to be found in [34]. Zardini considers a different sort of model, but defines consequence as depending upon two distinct standards for satisfaction, with the premises held to a higher standard than the conclusions, just as Smith and we do. Like our semantics,

[15] The only difference here occurs with *st*, because $Pa \wedge aI_P b \wedge bI_P c \models^{st} Pc$ (see Fact 3). The chaining stops here too, though, as $Pa \wedge aI_P b \wedge bI_P c \wedge cI_P d \not\models^{st} Pd$. Informally, *sc* and *ct* allow us to take only one step, and no more, along the similarity relation, while *st* allows us two steps, and no more. In all cases, though, it is the 'and no more' that blocks the threatening soritical reasoning.

[16] Some might argue that if a relation between sentence sets is nontransitive it is not a consequence relation. This seems to us merely a terminological issue.

and unlike Smith's, Zardini's semantics results in a nontransitive consequence relation. Unlike our semantics, however, Zardini's consequence relation is a weakening of the classical one. For example, the inference from ϕ, $\phi \rightarrow \psi$, and $\psi \rightarrow \chi$ to χ is invalid for Zardini, but it is *cc*-valid, and so *st*, *ct*, and *sc* valid as well.

When it comes to our weaker relations (*cs*, *tc*, and *ts*), they are no less weird in the full vocabulary than they are in the restricted vocabulary. *cs* and *tc* still don't obey uniform substitution, for one thing. *ts*, however, is now no longer the empty relation, since we have introduced vocabulary (identity and similarity) that does not differentiate between tolerant, classical and strict satisfaction. Thus, some inferences are now *ts*-valid. For example, $a = b \vDash^{ts} b = a$; $aI_P b, b = c \vDash^{ts} cI_P a$, $\forall x \forall y(x = y) \vDash^{ts} Pa \vee \neg Pa$, and so on. Of course, it follows that these inferences are then valid in all nine of our relations, as all are extensions of *ts*. In fact, Fact 4 holds, not just for $\vDash^{ss}$ and $\vDash^{tt}$, as it was stated, but for all nine notions of consequence, including *ts*.

3.5 Tableaux

Although it is widely recognized that using a non-transitive entailment relation might solve the sorites paradox (see e.g. [16, p. 20]), this is sometimes taken to be problematic. For example, Dummett [8] claims that transitivity is essential to any notion of proof. One way to meet that objection is to provide a proof system. Notice that it is standardly taken to be hard to give a proof theory for such a logic (see [15], for instance).

In fact, there is a single tableau system that is sound and complete for all nine notions of consequence discussed above, in the full vocabulary.

Definition 21 A *tagged sentence* is something of the form ϕ, m where ϕ is a sentence and m is either s, c, or t.

Definition 22 A T-model M *satisfies* a tagged sentence ϕ, m iff $M \vDash^{m} \phi$.

The nodes of our tableaux are tagged sentences rather than just sentences.[17] Depending on the main (and sometimes secondary) connective in a tagged sentence, we apply the appropriate rule to it to generate more tagged sentences. For familiar connectives ($\neg, \wedge, \forall, =$), we simply use ordinary tableau rules, and carry the tag along. A pair of examples:

$\phi \wedge \psi, s$	$\neg(\phi \wedge \psi), s$
ϕ, s ψ, s	$\neg\phi, s$ $\quad$ $\neg\psi, s$

[17] For details on how tableaux standardly work in first-order logic with identity, see e.g. [27] or [22]. Here, we assume familiarity with the general idea.

There are a few novel rules, however. The first batch covers the interaction between predication and similarity relations:

Pu, s uI_Pv, c	Pu, t uI_Pv, c
$P[v/u], c$ (for every such v)	$P[v/u], c$ (for a new v)

$\neg Pu, s$ uI_Pv, c	$\neg Pu, t$ uI_Pv, c
$\neg P[v/u], c$ (for every such v)	$\neg P[v/u], c$ (for a new v)

The second batch ensures that similarity and identity do not differ from tolerant to classical to strict:

$uI_Pv, s/t$	$\neg(uI_Pv), s/t$
uI_Pv, c	$\neg(uI_Pv), c$

$u = v, s/t$	$\neg(u = v), s/t$
$u = v, c$	$\neg(u = v), c$

(Here, s/t can be either s or t.) The third and final batch encodes the reflexivity and symmetry of the I_P relations:

.	uI_Pv, c
uI_Pu, c	vI_Pu, c

A branch *closes* iff it includes two tagged sentences of the form ϕ, c and $\neg\phi, c$. A tableau *closes* iff every branch on it closes.

Fact 5 (Soundness) *If a tableau built on a set* Γ *of tagged sentences closes, then there is no T-model M that satisfies every tagged sentence in* Γ.

Fact 6 (Completeness) *If a tableau built on a set* Γ *of tagged sentences does not close, then there is a T-model M that satisfies every tagged sentence in* Γ.

These facts hold both for our restricted vocabulary and for the full vocabulary. We omit the proofs here; they are simple modifications of usual soundness and completeness proofs for tableaux. (For an example of the usual proofs, see [22].) Note that these facts allow us to use our tableau for any of our nine notions of consequence:

Fact 7 $\Gamma \models^{mn} \Delta$ *iff a tableau built on* Λ *closes, where* $\Lambda = \{\gamma, m : \gamma \in \Gamma\} \cup \{\neg\delta, d(n) : \delta \in \Delta\}$.[18]

[18] As before, $d(c) = c$, $d(s) = t$, and $d(t) = s$.

We can also use these tableaux to explore still more articulated questions that do not reduce to questions about *mn*-consequence for any m, n. For any sets Γ, Λ, Σ of sentences, these tableaux will answer: is there any model where every sentence in Γ holds classically, every sentence in Λ holds tolerantly, and every sentence in Σ holds strictly?

3.6 The Sorites

How does this framework address the sorites paradox? In order to be as clear as possible, we consider two different versions of the sorites paradox. As above, we favor the relation *st*, and so we respond to the paradox from this perspective. Throughout, we assume a sorites series of objects $a_1, \ldots, a_n$ for the predicate P.

Version 1 One version of the sorites proceeds directly from similarity relations:

$$\frac{Pa_1 \quad \forall i \in [1, n-1](a_i I_P a_{i+1})}{Pa_n}$$

This version of the sorites is *st*-invalid. (There is no funny business with the quantification here; everything stays the same if we replace the quantified premise with n many atomic premises.) Of course, it is classically invalid as well; the interest in the *st* response comes not just from its invalidating this argument, but from invalidating this argument while *validating* each step. That is, the following argument is *st*-valid:

$$\frac{Pb \quad b\, I_P c}{Pc}$$

The system *st* can accomplish this balancing act because of its nontransitivity; transitive logics must validate both or neither of these inferences. We think, though, that the second inference is a good one, and that the first is not.

Version 2 A different version of the paradox has tolerance as a premise:

$$\frac{Pa_1 \quad \forall i \in [1, n-1](a_i I_P a_{i+1}) \quad \forall x \forall y (Px \wedge x I_P y \rightarrow Py)}{Pa_n}$$

(Again, we can replace the quantified premises with their instances without changing anything.) This version of the paradox is *st*-valid. After all, it is

classically valid, and, as we have seen, *st* is stronger than classical logic. Here, the reason we do not conclude that Pa_n holds, even tolerantly, is because there is an untrue premise. The third premise, tolerance, is not strictly true, and it is strict truth we require of our premises in *st*.

It is quite difficult for a conditional to be strictly true. Remember, we understand $\rightarrow$ as a material conditional: $\phi \rightarrow \psi$ is to be read as $\neg\phi \vee \psi$. This allows even $\phi \rightarrow \phi$ to fail when ϕ is a borderline sentence; we should not be surprised that tolerance does not meet this high a standard.

However, as we have mentioned before, tolerance does meet the lower standard of tolerant validity. This version of the sorites paradox reminds us that one must be careful using even valid sentences as premises, if the standard for validity differs from the standard that premises must meet.

It may at first seem to be in tension with our tolerance-preserving approach to call this argument valid, and refuse to strictly assent to tolerance, but we think that once an appropriate pragmatic framework is in place, the appearance of tension dissipates. We turn to this issue in Section 4.1.

4 The Pragmatics of Vague Predicates

In this section we discuss some applications of our framework to the semantics and pragmatics of vague predicates. Our framework allows us to define two dual notions of interpretation for a predicate on top of the classical one, namely strict and tolerant. This raises the question of which of these interpretations is likely to be preferred in interpreting and using vague predicates. We think that both strict and tolerant interpretations are needed in an empirically adequate treatment of vague predicates.

This obligates us to give some account of when each sort of interpretation plays a role. When someone asserts a vague sentence, are we to interpret it strictly, classically, or tolerantly? We appeal to an independently motivated pragmatic mechanism (the "strongest meaning hypothesis") that delivers either strict or tolerant interpretations, depending on the context. In this our treatment of assertion agrees with the pragmatic account of vague predicates recently proposed by Alxatib and Pelletier in [1]. In the second part of this section, we relate our framework to the experimental data they obtained, as well as to the earlier findings reported by Ripley in [23] and by Serchuk, Hargreaves and Zach in [25].

4.1 Meaning Strengthening

To get a sense of the mechanism we postulate, it might help to consider a particular objection to our semantic framework. Consider Fred and Bert, two borderline cases of 'tall'. Suppose that Bert is slightly taller than Fred. In this scenario, it seems clear that it is inappropriate to assert 'Fred is tall and Bert is not tall'. Nonetheless, our framework allows this sentence to be tolerantly (albeit not strictly) satisfied. This is at least prima facie counterintuitive.

In this section, we explain why it is *pragmatically inappropriate* to assert 'Fred is tall and Bert is not tall' in the above circumstances, even though the sentence might be (tolerantly) true. The explanation will be that without any further information, a hearer of this utterance will conclude from this that Fred is significantly taller than Bert, which is false. To account for this reasoning we make use of a pragmatic theory of preferred interpretation, in particular, an interpretation strategy known as *the strongest meaning hypothesis*.

A theory of preferred interpretation is crucial to determine what was meant by the use of a particular sentence that is semantically ambiguous. Consider, for instance, the case of pronoun resolution. Look at the following simple discourse (2).

(2) John met Bill at the station. *He* greeted *him*.

The pronouns *he* and *him* could refer to either John or Bill. Still, there is a preference (for reasons of syntactic parallelism) for interpreting *he* as John and *him* as Bill, and in 'normal' circumstances, this is the way we proceed. But this preference can be overruled if we add additional information. For instance, if we add 'John greeted him back', we have to reinterpret *he* as Bill, and *him* as John, due to the indefeasible semantics associated with the adverb *back* (cf. [12]). Thus, if sentences are semantically ambiguous, it might still be that one interpretation is more preferred than others, and in normal circumstances this is the way we actually interpret the sentence. This preferred interpretation might be overruled, however. Asher and Lascarides [2] observe a similar pattern with temporal anaphora: normally the event a first sentence in simple past is about is temporally located before the event the consecutive sentence with simple past is about. But world-knowledge sometimes forces us to interpret otherwise, as in the discourse 'John fell. Mary pushed him'.

There might be various reasons why, out of context, one interpretation of a sentence is preferred to another one. In the examples discussed above, the preference was due to syntactic and pragmatic factors, respectively. In some interesting cases, however, the preference is due solely to *semantic* factors. Take, for instance, the interpretation of plural reciprocals. It is well-known that sentences like 'The children followed each other' allow for many different interpretations. Still, such sentences are most of the time understood pretty well: each child followed another child. Dalrymple et al. [6] propose that this is due to a particular interpretation strategy. According to their "Strongest meaning hypothesis" a sentence should preferentially be interpreted in the *semantically* strongest possible way. This simple strategy predicts surprisingly well, and has become popular to account for other phenomena too (cf. [33]). But it is important to note here that the hypothesis used is one of *preferred* interpretation only: adding more information might make a stronger interpretation impossible. If we add 'into the church', for instance, our original sentence has to be re-interpreted, and can at most mean that any child followed, *or was followed*, by another child.

Observe that any sentence that involves a predicate like 'tall' is according to our analysis semantically ambiguous as well, or at least allows for different

semantic interpretations. The reason for this, of course, is that such sentences can be interpreted strictly, classically, and tolerantly, and these interpretations are typically different. We have seen above that for each sentence ϕ it holds that $[\![\phi]\!]^s \subseteq [\![\phi]\!]^c \subseteq [\![\phi]\!]^t$. If we adopt the strongest meaning hypothesis, this means that sentences involving vague predicates are preferably interpreted strictly, and that the tolerant interpretation is less preferred than the classical one.

Consider the sentence 'Fred is tall and Bert is not tall' again. If we interpret this sentence in the preferred strongest possible way, it can only be true if Fred is strictly tall and Bert is strictly not tall, which implies that Fred must be significantly taller than Bert. Thus, out of context it is inappropriate to assert that 'Fred is tall and Bert is not tall' if Fred is similarly tall to Bert. This explains why the sentence is neither appropriately asserted, nor interpreted as true, in the circumstances that make it only tolerantly true. Other sentences, however, can only be tolerantly true, and it is thus predicted that such a sentence is interpreted in this tolerant way. This holds, obviously, for sentences like 'Bert is tall and Bert is not tall'. Since this sentence cannot ever be strictly or classically true, it would be odd to interpret it strictly or classically; a tolerant interpretation is called for. On the tolerant interpretation, this sentence expresses the claim that Bert is a borderline case of 'tall'. This, it seems to us, is in order. (Also, that this sentence cannot be strictly (or classically) true might help explain why [10, 14], and others feel that this sentence cannot be true; they do not consider tolerant truth.)

Though this is appealing, there still might seem to be a problem with our explanation: if it is known in the context of interpretation that Bert is only slightly taller than Fred, doesn't this mean that 'Fred is tall and Bert is not tall' should be interpreted tolerantly after all, and thus taken to be true? We don't think so; it would still be inappropriate to assert that Fred is tall and Bert is not tall, because there is an alternative sentence that the speaker could have uttered that could be interpreted in a stronger way and still be true (e.g. 'Bert is tall and Fred is not tall'). This type of reasoning is both natural and very much in the Gricean spirit.

This pragmatic strategy towards assertions has another pleasant consequence: we predict that 'Bert is tall or Bert is not tall' is preferably interpreted strictly, which means that it is not counted as automatically true, and interpreted as an informative statement.

This strategy also provides the explanation we promised in Section 3.6. Recall that the version of the sorites paradox that includes tolerance as a premise is *st*-valid; we claim it is unsound because the tolerance premise is not strictly true. This might at first seem implausible because tolerance *seems* true, but our pragmatic hypotheses here can explain this seeming. Given the truth (even just the tolerant truth!) of the first premise, Pa_1, and the truth (even just the tolerant truth!) of $\neg Pa_n$, where Pa_n is the sorites's implausible conclusion, tolerance cannot be strictly satisfied; it can only be tolerantly satisfied. Because of this, it would be uncooperative to interpret it strictly. Thus, when we are faced with the *st*-valid version of the sorites argument, tolerance seems true to

us, but this is because pragmatic mechanisms lead us to interpret it tolerantly instead of strictly.

4.2 Psycholinguistic Evidence

To substantiate our discussion, we confront our treatment of penumbral connections with some recent psycholinguistic data on the semantic treatment of vague predicates established independently by [1, 23] and [25]. Overall, the data suggest that subjects do not preserve classical logical truths for borderline cases. Rather they appear to reason either tolerantly, or strictly, but in agreement with the strongest meaning hypothesis.

4.2.1 Contradictions and Borderline Cases

Ripley tested subjects' level of agreement to various sentences involving the vague predicate "near". Subjects were shown a figure representing seven pairs (A to G) each consisting of a square and a circle at decreasing distances from each other. Pair A was a clear non-case of "near", and Pair G was a clear case of "near". Subjects were asked for each pair to rate their agreement to one of several syntactic variants of the sentence "the circle is near the square and it isn't near the square", including "the circle neither is nor isn't near the square". What he found is that a significant proportion of subjects fully agree with these sentences in some cases, and moreover that agreement is significantly higher for the median pair C (in which the circle is about half way between what it is in the extreme pairs A and G).

Similarly, Alxatib and Pelletier showed subjects a picture representing five men of different heights, with an explicit indication of their actual heights, in order to test for people's understanding of the vague predicate "tall". Subjects were then asked to respond to four questions for each man in the drawing, namely to judge whether it is true or false that the man is *tall*, *not tall*, *tall and not tall*, and finally *neither tall nor not tall* (with the possibility to give a third answer: 'Can't tell'). What Alxatib and Pelletier found is that for the man #2 of median size on the figure, namely of size 5'11", 44.7% of the subjects responded True to 'X is tall and not tall', and 53.9% likewise responded True to 'X is neither tall nor not tall'. What is significant for our purpose is that this proportion of True answers to classical contradictions was again significantly higher for this man than for men of other sizes in the series. Moreover, more than half of the subjects who judged #2 both tall and not tall judged #2 neither tall nor not tall (64.7%), and conversely (53.7%).

Overall, the results obtained by Ripley as well as Alxatib and Pelletier indicate that the subjects' level of agreement to contradictions is therefore significantly higher for borderline or intermediate cases than for the extreme cases in their displays. Prima facie, these data are therefore consistent with the view that sentences of the form $Pa \wedge \neg Pa$ can be used tolerantly for borderline cases. Moreover, they indicate that the predictions of either subvaluationism or supervaluationism for borderline cases are not empirically adequate (see [1, 23] for discussions).

Serchuk et al. [25] used a different methodology to test semantic judgments about borderline cases. They did not confront subjects with actual stimuli, but rather gave them a linguistic scenario in which a character named Susan was described as "somewhere between women who are clearly rich and women who are clearly non-rich"; a similar kind of scenario was offered for the adjective "heavy". They asked subjects to evaluate various sentences, including "Susan is rich and Susan is not rich" (and similarly for "heavy"). Their answer space was larger than that in [1], as they gave their participants the choice between True, False, Both, Neither, Partially True, and Don't Know. For that particular sentence type they found that more than 55% declared the sentence False, against about 19% judging it True (with lower ratios in each of the other answers). They conclude from their experiment that subjects "tend to preserve the law of contradiction" for borderline cases. Due to the larger answer space, it is hard to compare their results with Ripley's or Alxatib and Pelletier's however. Methodological differences between the two experiments might explain the different results as well—in particular subjects might be more willing to preserve the law of non-contradiction when they issue judgments based on an abstract representation of a borderline case, rather than when they are driven by the actual perception of a borderline case. In any case, however, their results for disjunction and the law of excluded middle (see below) confirms the hypothesis that subjects do not reason purely classically for borderline cases, in contradistinction to the predictions of sub- or super-valuationism.

4.2.2 Alxatib and Pelletier on Meaning Strengthening

One of the striking results of Alxatib and Pelletier is that very few of those who check True to "both tall and not tall" for the man of intermediate size also check True to "tall" and to "not tall" separately (only 2.9%); by contrast, 32.4% of the same subjects who assented to "both tall and not tall" checked False to the conjuncts "tall" and "not tall" separately.

In their paper Alxatib and Pelletier proposed a pragmatic explanation for this phenomenon that antedates our account on two aspects. First of all, Alxatib and Pelletier propose:

> "an assumption that may seem somewhat controversial: that a given vague predicate has two possible interpretations, a super-interpretation and a sub-interpretation, in the same way that a vague expression containing negation can be interpreted strongly (i.e. super-interpreted), or weakly (i.e. sub-interpreted)."

Alxatib and Pelletier do not specify an explicit compositional semantics for these notions in their paper. However, the distinction they make between a sub-interpretation and a super-interpretation can be captured exactly in terms of our distinction between a tolerant (for their "sub-") and a strict interpretation (for their "super-") for predicates. The distinction Alxatib and Pelletier make between super-interpretation and sub-interpretation of the

negated predicate "not tall" is indeed presented in their paper as a distinction between two kinds of negation, but can be viewed equivalently as a scope distinction relative to a silent operator. Thus, the super-interpretation in this case corresponds to "definitely not tall", while the sub-interpretation corresponds to "not definitely tall". Though we did not introduce a "definitely" operator in our language, note that we could in principle introduce an operator $\Box$ such that $M \models^c \Box Px$ iff $M \models^s Px$ (see [29], where this is done). Consequently, $M \models^s \neg Px$ would mean that x is definitely not P, while $M \not\models^s Px$, or equivalently $M \models^t Px$, would mean that x is not definitely tall.

The second element of their account closely corresponds to the strongest meaning hypothesis we formulated in the previous section, and is indeed viewed as a particular case of it by Alxatib and Pelletier (see [1], footnote 20), namely:

> Of the two interpretations, the super- and the sub-, the maxim of quantity demands that the stronger of the two be intended. If *a* is of borderline height, the statement is likely to be disagreed with, since *a* does not qualify as super-tall, or super-not-tall.

The upshot is that subjects who check True to "both tall and not tall" for borderline cases interpret the whole sentence tolerantly, though the same subjects who check False to "Tall" and "Not Tall" respectively interpret each of the latter strictly. Each of these is compatible with the strongest meaning hypothesis.

In the previous section, we mentioned that in the same way in which classical contradictions of the form "P and not P" would not necessarily be false of borderline cases, classical tautologies of the form "P or not P" could be false. Neither Ripley nor Alxatib and Pelletier tested for such disjunctions directly, but only for "neither tall nor not tall". In the borderline case, the data suggest that a large proportion of subjects understand this strictly.[19] However, Serchuk et al. tested disjunctive sentences of the form "Either Susan is rich or Susan is not rich" in their scenario and found a large proportion of False answers (39%). They do not report about the behavior of those subjects who check False regarding each of the disjuncts; however, their data indicate a lower global proportion of True answers to the sentence "Susan is not rich" (21%). Overall, their finding is therefore compatible with the idea that each of the disjuncts is interpreted strictly in this case, as is the disjunction itself, again in compliance with the strongest meaning hypothesis.[20]

[19] We can imagine either that subjects understand the sentence as "(strictly) neither tall nor not tall", or that they understand it as "neither (strictly) tall nor (strictly) not tall". We set aside discussion of the choice between these two understandings.

[20] Thus Serchuk and colleagues present their data as "consistent with Keefe's confusion hypothesis", namely the view that speakers "confuse" sentences of the form "*Fa* or not *Fa*" with "definitely *Fa* or definitely not *Fa*" (see [16]). In our view, talk of meaning strengthening is more appropriate to describe this phenomenon than talk of a confusion about meaning.

4.2.3 Borderline Cases and Similarity

A final question we may ask concerns the psychological plausibility of our similarity based semantics regarding the attribution of vague predicates. If we think of Alxatib and Pelletier's experiment, we could ask why it is for the man of median size that the proportion of "neither tall nor not tall" and "both tall and not tall" answers is the greatest. Similarly, in the case of Ripley's experiment, it is for the pair at roughly median distance between the extreme pairs that the "neither near nor not near" and "both near and not near" answers peak.

Several explanations of this phenomenon are conceivable. One is that more or less equidistant cases between focal points (conceived as prototypes) are part and parcel of what it means to be a borderline case (see [7]). Similarly here, borderline cases between P and not P can be viewed as cases equisimilar to P and not P cases. Van Rooij [28] suggests that when we have to judge whether "x is tall" or not relative to a comparison class, we first delineate between the tallest and between the shortest, so as to leave a gap. In the case of Alxatib and Pelletier's design, men #3 and #5 are visibly the tallest and of roughly equal size, and men #1 and #4 are visibly the shortest, and of roughly equal size. Furthermore, these two sets are sufficiently separated, namely the tallest of the short (#4) is sufficiently far from the shortest of the tall (#5). However, #2 is hard to assign to either "tall" or "not tall", precisely because #2 is roughly equally close to #4 and to #5 in size.

In model theoretic terms, letting P stand for "tall" we can represent the situation by a model in which we have the non-transitive chain of pairwise similarities: $1 \sim_P 4 \sim_P 2 \sim_P 5 \sim_P 3$. Suppose subjects first include 1 and 4 in the classical extension of P, and 5 and 3 to the extension of $\neg P$. Then, irrespective of whether 2 is assigned to P or not, the resulting total model is one in which 2 is predicted to be neither strictly tall nor strictly not tall; that is, tolerantly tall and tolerantly not tall. In our presentation of tolerant and strict semantics, we assumed total models from the start. However, the present example suggests that we could propose an alternative formulation of the semantics starting from partial models, and yet remain faithful to the definition of borderline cases as cases similar to both P and not P cases. We leave the details of this more refined approach for subsequent work.

5 Conclusion

In this paper we proposed a new semantic framework for the treatment of vague predicates. As we discussed along the way, our framework shares a number of features with extant semantics for vagueness, though we believe it differs from each in some important respects.

First of all, our similarity-based semantics for first-order logic rests on the idea that vagueness is tied in an essential way to non-transitivity, whether of indifference or indiscriminability. In this, the framework agrees in particular with

one of the central hypotheses of Williamson's epistemic theory of vagueness (see [32]), but it does not commit us to the view that the classical extensions we stipulated in T-models necessarily reflect the "objective" meaning of vague predicates. Rather, on our approach these models can be used to describe the internal representations of subjects confronted with the task of categorizing objects based on how similar they look.

Secondly, we have argued that the duality between the notions of tolerant and strict truth gives a natural characterization of borderline cases. It also allows us to validate the tolerance principle; finally, the definition of a mixed relation of logical consequence, from strict to tolerant, allows us to preserve most of the classical rules of inference, in particular modus ponens, only at the expense of having a non-transitive notion of logical consequence. In our opinion, the loss of transitivity for logical consequence is less dramatic a cost than the loss of modus ponens in response to the sorites paradox.

Finally, we have seen that the very duality of strict and tolerant interpretations can be seen to match the experimental data recently established regarding how people evaluate complex sentences containing borderline cases.

Several issues remain to be investigated. As mentioned in the last section, one aspect we did not go into concerns the possibility of presenting a tolerant/strict semantics based on partial rather than total interpretations. Secondly, in this paper we mostly focused on the sorites paradox and on the status of borderline cases. We deliberately set aside the issue of higher-order vagueness, in particular because it concerns the behavior of an operator like "definitely" or "clearly" in a richer language. As briefly pointed out, it would be very natural to introduce a "definitely" operator to mirror the notion of strict truth syntactically (see [29]).

Acknowledgements We thank two anonymous reviewers for detailed and helpful comments. Further thanks go to various colleagues and audiences for their valuable feedback at conferences and lectures held in Kolkata, Breclav, Geneva, Trento, Barcelona, Nancy, Amsterdam, Paris, Pittsburgh, Aberdeen and St Andrews, where different parts and stages of this paper were presented between September 2009 and June 2010. We wish to thank in particular Mihir Chakraborty, Philippe de Groote, Floris Roelofsen and Nicholas Asher for very helpful remarks and suggestions. Special thanks go to Sam Alxatib, Jeff Pelletier, Phil Serchuk, and Elia Zardini for valuable input and exchanges based on each of their recent works on vagueness. This work was done with main support from the Agence Nationale de la Recherche, program 'Cognitive Origins of Vagueness', grant ANR-07-JCJC-0070, as well as the ESF program 'Vagueness, Approximation and Granularity', the NWO project 'On vagueness—and how to be precise enough', and the project 'Borderlineness and Tolerance' (Ministerio de Ciencia e Innovación, Government of Spain, FFI2010-16984), all of which are gratefully acknowledged. We also thank the Formal Epistemology Project of the University of Leuven and particularly I. Douven and R. Dietz, with whom the workshop 'Vagueness and Similarity' held in Paris in May 2010 was coorganized.

References

1. Alxatib, S., & Pelletier, J. (2010). The psychology of vagueness: Borderline cases and contradictions. *Mind and Language*. (forthcoming).

2. Asher, N., & Lascarides, A. (1993). Temporal interpretation, discourse relations, and commonsense entailment. *Linguistics and Philosophy, 16*, 437–493.
3. Beall, J., & van Fraassen, B. C. (2003). *Possibilities and paradox: An introduction to modal and many-valued logic*. Oxford: Oxford University Press.
4. Cobreros, P. (2008). Supervaluationism and logical consequence: A third way. *Studia Logica, 90*(3), 219–312.
5. Cobreros, P. (2010). Varzi on supervaluationism and logical consequence. *Mind*. (forthcoming).
6. Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., & Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy, 21*, 159–210.
7. Douven, I., Decock, L., Dietz, R., & Egré, P. (2010). Vagueness: A conceptual spaces approach. (manuscript, under review).
8. Dummett, M. (1975). Wang's paradox. *Synthese, 30*, 301–324.
9. Egré, P., & Bonnay, D. (2010). Vagueness, uncertainty, and degrees of clarity. *Synthese, 174*, 47–78.
10. Fine, K. (1975). Vagueness, truth, and logic. *Synthese, 30*, 265–300.
11. Goodman, N. (1951). *The structure of appearance*. Cambridge: Harvard University Press.
12. Grosz, B., Joshi, A., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21*, 203–226.
13. Hyde, D. (1997). From heaps and gaps to heaps of gluts. *Mind, 106*, 641–660.
14. Kamp, H. (1976). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics for natural language*. Cambridge: Cambridge University Press.
15. Kamp, H. (1981). The paradox of the heap. In U. Mönnich (Ed.), *Aspects of philosophical logic*. D. Reidel.
16. Keefe, R. (2000). *Theories of vagueness*. Cambridge: Cambridge University Press.
17. Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica, 24*, 178–191.
18. Mendelson, B. (1975). *Introduction to topology* (3rd ed.). Dover reedition 1990. New York: Dover.
19. Pawlak, Z. (2005). A treatise on rough sets. In J. F. Peters, & A. Skowron (Eds.), *Transactions on rough sets IV* (pp. 1–17). Berlin: Springer.
20. Pinkal, M. (1995). *Logic and Lexicon*. Dordrecht: Kluwer Academic Publishers.
21. Priest, G. (1979). Logic of paradox. *Journal of Philosophical Logic, 8*, 219–241.
22. Priest, G. (2008). *An Introduction to non-classical logic: From if to is* (2nd ed.). Cambridge: Cambridge University Press.
23. Ripley, D. (2009). Contradictions at the borders. In R. Nouwen, R. van Rooij, H.-C. Schmitz, & U. Sauerland (Eds.), *Vagueness in communication*. Berlin: LICS, Springer. (forthcoming).
24. Ripley, D. (2010). Sorting out the sorites. In F. Berto, E. Mares, & K. Tanaka (Eds.), *Paraconsistent Logic (tentative title)*. (forthcoming).
25. Serchuk, P., Hargreaves, I., & Zach, R. (2010). Vagueness, logic and use: Four experimental studies on vagueness. *Mind and Language*. (forthcoming).
26. Smith, N. J. J. (2008). *Vagueness and degrees of truth*. Oxford: Oxford University Press.
27. Smullyan, R. M. (1995). *First-order Logic*. New York: Dover.
28. van Rooij, R. (2010a). Implicit vs. explicit comparatives. In P. Egré, & N. Klinedinst (Eds.), *Vagueness and language use*. Palgrave Macmillan.
29. van Rooij, R. (2010b). *Vagueness, tolerance, and non-transitive entailment*. (manuscript).
30. Varzi, A. (2007). Supervaluationism and its logics. *Mind, 116*, 633–676.
31. Weber, Z. (2010). A paraconsistent model of vagueness. *Mind*. (forthcoming).
32. Williamson, T. (1994). *Vagueness*. London: Routledge.
33. Winter, Y. (2001). Plural predication and the strongest meaning hypothesis. *Journal of Semantics, 18*, 333–365.
34. Zardini, E. (2008). A model of tolerance. *Studia Logica, 90*, 337–368.